INTRODUCTION TO STATISTICAL MATHEMATICS

INTRODUCTION TO

Statistical Mathematics

A. M. MATHAI

INTRODUCTION TO
STATISTICAL MATHEMATICS

INTRODUCTION TO STATISTICAL MATHEMATICS

(Mathematics of Stochastic Variables)

BY

A. M. MATHAI

Assist. Professor of Mathematics
McGill University

1967

S. CHAND & CO.

DELHI — NEW DELHI — JULLUNDUR
LUCKNOW—BOMBAY—CALCUTTA—MADRAS

S. CHAND & CO.

Ram Nagar —	NEW DELHI
Fountain —	DELHI
Mai Hiran Gate —	JULLUNDUR
Aminabad Park —	LUCKNOW
167, Lamington Road —	BOMBAY
32, Ganesh Chandra Ave. —	CALCUTTA
35, Mount Road —	MADRAS

Price : **Rs. 20.00**
U.S. $ 8·00

Published by S. Chand & Co., Ram Nagar, New Delhi-1
and printed at Rajendra Printers, Ram Nagar, New Delhi-1

To

Rev. Fr. Joseph Kureethadom

whose constant encouragements have contributed a great deal towards my academic achievements

PREFACE

This is an introductory book on Mathematics of stochastic variables. The book deals with elementary Probability and Statistics. This can be used as a book for self study or for a one year slow course in Probability and Statistics. Since it is also intended as a book for self study, the various symbols and letters used, are explained then and there throughout the book. The pre-requisite is one year Calculus. But a person with high school Mathematics can follow the book except the few sections which use Calculus extensively. In Chapter 1 an introduction to Set Theory and Linear Algebra is given. The pre-requisite for this chapter is only high school Mathematics. The later sections of Linear Algebra may be omitted because they are not very much used. These facts are mentioned in the different sections. An attempt is made to make the book semi-rigorous.

It is a fairly balanced treatment of theory and applications and this will give a sufficiently good background for further studies. The book is based on the topics covered in an introductory course in Statistics for the General Arts and Science students at McGill University.

The book is intended for :

(1) self study.

(2) a two semester course in Probability and Statistics.

(3) a one semester course in Probability (Chapters 2-5).

(4) a one year course in Indian universities.

Special features :

1. A new approach—built up on stochastic variables and the operator called 'Mathematical Expectation'; uniformity in notations.
2. Student's difficulty of distinguishing discrete, continuous, finite, infinite, observed and hypothetical populations, is avoided by properly defining and developing the theory, based on the theory of sets.
3. The dialogue is designed to suit the particular age group of students who are likely to take the course.
4. A number of worked examples are given in each section and a good number of examples are taken from problems of day to day life.
5. The development of the theory is very slow in the beginning chapters and the discussion is precise and a minimum in later chapters.
6. An insight into the advanced and various related topics, is given to the reader in every section.
7. Summary of correspondence between topics, important results, formulae etc. is given in every chapter.

8. A good number of problems, among which some of them supplement the theory already discussed, are given at the end of each section.
9. A unified theory of statistical inference is developed in the last chapter.
10. Consistency is kept throughout ; back references, references to the later sections and repetitions are kept a minimum.
11. Trivial results and results stated without proof are given in the form of comments after illustrating them by examples.
12. Every subsection is illustrated by atleast one worked example.

A list of notations is given at the end of the book. In numbering sections, equations and problems the following notations are used. For example,

Problem 10·24 (24th problem in chapter 10).

Section 2.3.2 (second subsection of the 3rd section of chapter 2).

Equation (8.21) (21st equation in chapter 8).

Answers of even numbered question in the whole book and of almost all questions in chapters 1—10 are given at the end of the book. Answers of almost all even numbered questions are rechecked.

Several people have spent their precious time in helping the author to complete the book. The author wishes to express his sincere thanks to the following professors, Dr. Mir M. Ali, Dr. J.C. Ahuja, Dr. V. Seshadri, Dr. D. Dawson, Dr. B.D. Aggarwala, for their valuable comments on some sections and Dr. M. Stephens, Dr. M. Csorgo, Dr. E. Saleh, Dr. G.P. Patil, Dr. J.K. Wani, Dr. R.K Saxena for the interesting discussions on some topics. The author extends his heartfelt thanks to prof. T.D. Dwivedi for helping the author to proofread the materials and for checking the answers of some exercises. The author would like to thank Dr. P.P Singh, Dr. N.K. Mathur and Shri S.K. Agrawal for their help in making a detailed index, Miss H. Schroeder for taking up the hard job of typing the first draft of the manuscript, Mr. M. Yalovsky and Mrs. F. Gordon for their comments, the National Research Council of Canada, Prof E.M. Rosenthall and the Department of Mathematics, McGill University for the financial assistance in computing a few tables.

The author is indebted to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., Cambridge, and to Oliber and Boyd Ltd. Edinburgh, for their permission to reprint Table No. 5 from their book 'Statistical Methods for Research Workers' and to Cambridge University Press, Prof. D.V. Lindley and Prof. J.C.P. Miller for their permission to reprint Tables 6, 7, 8 and 9.

A.M. Mathai

CONTENTS

LIST OF TABLES

SYMBOLS

CHAPTER 1

INTRODUCTION

1·0. Introduction. In this chapter some of the basic ideas which are required for the developments in later chapters are discussed. A reader who is familiar with the definitions of set, population, sample, vector and matrix may omit this chapter. Mainly the definitions and elementary properties of sets, populations, vectors and matrices are considered. Most of these results are utilized in later chapters.

1·1 SETS

1·11. Definition. A collection or an aggregate of well-defined objects is called a set. These objects which belong to the set are called elements of the set. Sets are usually denoted by capital letters A, B, C etc., and their elements by small letters a, b, c etc. Set, aggregate, group, collection etc., are synonyms in ordinary language but in mathematical language they have different meanings. 'well-defined' here means that any object may be classified as either belonging to the set A or not belonging to the set A.

Notations. $a \in A$ means that a is an element of the set A where $\in$ is a Greek letter 'epsilon'.

$b \notin A$ means that b is not an element of the set A.

$A=\{0, 1, -100\}$ means that A is a set with elements 0, 1 and -100. $0 \in A$, $1 \in A$, and $-100 \in A$ but for example $20 \notin A$.

$B=\{x \mid -10 \leqslant x \leqslant 25\}$ means that B is a set of all points which lie between -10 and 25 (including end points). This defines an interval (closed), $[-10, 25]$.

$C=\{(x, y) \mid 2x+3y=5\}$ means that the set of all paired values (x, y) for which the equation $2x+3y=5$ is satisfied.

$D=\{(a, b) \mid a, b \in A\}$ means the set of all pairs of elements (a, b) where a and b are elements of a set A.

Ex. 1.11.1. *The set of numbers between 1 and 3.*

Comments. This set contains an infinite number of elements since there are an infinite number of numbers between 1 and 3.

Ex. 1.11.2. *The set of definitions in a particular text book.*

Comments. This set contains a finite number of elements. It may be noticed that the elements of a set need not be numbers and that a set may contain a finite or an infinite number of elements.

Ex. 1.11.3. *The set of assumptions for a particular mathematical statement.*

Comments. The elements of a set can be real or abstract quantities, animate or inanimate objects etc.

Ex. 1.11.4. *The set of books, pens and students in a class room.*

Comments. A set may contain different types of objects or objects having different characteristics. If Mr. Fox is a student in the class he is an element of the set. That is, $a \in A$ where a is Mr. Fox and A is the set of books, pens and students under consideration. The reader is advised to construct some examples.

1.12. Sets and Populations. A set whose every element can be characterized by K characteristics may be called a K-variate population, where K is a number greater than or equal to one ($\geqslant 1$). If $K=1$ the set is a univariate population ; if $K=2$ it is a bivariate population, etc. We will start with this definition for a population and as we proceed further we will discuss populations defined by a given set and populations defined by a stochastic variable, etc. It may be noticed that this notion of population is not always the same as the notion of population used in ordinary conversation.

Ex. 1.12.1. *The measurements of heights of all the students in a university at a particular time.*

Comments. This is a finite population because the set contains only a finite number of elements. Every element is characterized by a characteristic, namely, 'height measurement' of a student. This is an example for a one variate or univariate population.

Ex. 1.12.2. *A set of simultaneous measurements of heights, weights and lengths of right arm of all the citizens in a particular city at a particular time.*

Comments. Here every element of the set is a collection of three numbers (height, weight and length of right arm) or each

element is specified by three characteristics. This is an example of a trivariate or three variate finite population.

Ex. 1.12.3. *The set of true effects of a particular drug on all the animals of a particular category.*

Comments. This may be considered as a univariate but hypothetical population. According to the definition it may be noticed that a population may be finite or infinite, real or hypothetical, discrete (individually distinct) or continuous. The reader may construct three examples each of (a) univariate population (b) bivariate population (c) sets which cannot be considered as populations at all.

1.13. Outcome Set. The set whose elements are all possible outcomes of an experiment, is called an outcome set. This outcome set is also called sample space, possibility space, universal set, sure event etc. Thus an outcome set is a particular type of set, where the elements are the possible outcomes of an experiment. An experiment may be defined as a procedure which results in some outcomes in a particular situation. An outcome is a single realization of a phenomenon under consideration, under the assumptions and notations used for the procedure (experiment). The outcomes need not always be numbers or quantities which are representable in terms of numbers. A philosophical discussion of experiment and outcomes is not attempted here.

The outcomes of an experiment may be represented in different ways in other words the possible geometrical or algebraic or other representation of the outcome set is not unique. That representation where the elements do not represent more than one distinct outcome in some sense, is usually taken as the outcome set. Thus an outcome set does not allow any subdivision of its elements.

Ex. 1.13.1. *Consider an experiment of throwing a coin twice. Let one side (say, head) be denoted by 1 and the other side (say, tail) be denoted by 0. If we rule out the possibility of the coin standing on its edge, then the possible outcomes are* $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ *where the first element in a bracket denotes the result on the first trial. Here the outcome set is the set of 4 ordered pair of numbers given above.*

Comments. These may be represented as four geometrical points in a two dimensional space. Any single outcome here may be represented by a pair of numbers (a, b) where a and b take values 0 and 1 or where a and b are defined on the set $\{0, 1\}$. That is, $S=\{(a, b) \mid a, b \in \{0, 1\}\}$ where S denotes the outcome set. According to the above assumptions and terminology the outcome sets when a coin is thrown once, twice,..., n times, are given as follows :

No. of times the coin is thrown	*Possible outcomes*	*Total number of outcomes in the outcome set*	*The outcome set*
Once	0, 1	2	$\{a\}$ where a is defined on the set $\{0, 1\}$ or $\{a \mid a \in \{0, 1\}\}$
Twice	(0, 0), (0, 1), (1, 0) and (1, 1).	$2^2=4$	$\{(a, b) \mid a, b \in \{0, 1\}\}$
Thrice	(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1),	$2^3=8$	$\{(a. b, c) \mid a, b, c \in \{0, 1\}\}$
⋮	⋮	⋮	⋮
n-times	(0, 0,..., 0),...	2^n	$\{(a_1, a_2, \ldots, a_n) \mid a_1, \ldots, a_n \in \{0, 1\}\}$

If we consider the geometrical representation of the outcomes when the coin is tossed n times, we get 2^n points in an n-dimensional space. (a_1 is read as a one etc). The reader may evaluate the outcome sets when a die is thrown once, twice, etc. (a die is a cube with the six faces marked with the numbers 1, 2, 3, 4, 5, 6).

1.14. Subsets. A set B is said to be a subset of a set A if all the elements of B are also elements of A. This relationship is denoted as $B \subset A$ (B is contained in A or A contains B).

Ex. 1.14.1. *B—the set of male students in a class*

A—the set of students in the same class.

Comments. If there are no male students in the class, then B has no elements. Such a set is called a null set and is usually denoted by ϕ (Greek letter called Phi). So evidently $\phi \subset A$ where A is any set. If all the students are male students then B is the same as A. Thus equality of two sets A and B may be defined as $A=B$ if $A \subset B$ and $B \subset A$. It may also be noticed that $A \subset A$ or any set is a subset of itself.

Ex. 1.14.2. $B=\{x \mid 1 \leqslant x \leqslant 3\}$ *or the set of numbers between 1 and 3 (both 1 and 3 inclusive).*

$A=\{x \mid 1 \leqslant x \leqslant 5\}$ *or the set of numbers between 1 and 5 (both 1 and 5 inclusive).*

Comments. Evidently $B \subset A$. But it may be noticed that both A and B are infinite sets or sets containing infinite number of elements.

1.15. Subsets and Samples. If A is a population and if $B \subset A$ then B is called a sample with respect to A. According to

the definition B itself is a population if A and B are not considered together. If the formulation of the results or inference from B can be extended to A in some sense then B is called a representative sample from A.

Ex. 1.15.1. *The heights of students in a particular class is a sample of heights of students in the university to which the class belongs.*

Comments. This is an example of a univariate population and a sample from it. If in this sample 20% of the students have heights over 5′ 6″ then the statement that 20% of the students in the university have heights over 5′ 6″ can not be made unless the sample is representative of the population. In other words we are not justified in making such a statement based on the sample unless the sample is selected in such a way that such statements are justifiable in some sense. This aspect will be discussed in more detail in the chapter on sampling.

Ex. 1.15.2. *The outcome of rolling 2 dice three times is a sample from the hypothetical population of all possible outcomes of throwing 2 dice.*

Comments. This is an example of a sample from a multivariate population.

Ex. 1.15.3. *The set of birds captured from a particular place by an experimental scientist is a sample of birds at that place at that time.*

Comments. Generalization of his findings based on these birds can not be done unless the captured birds form a representative sample.

1.16. Events. A subset A of an outcome set S is called an event. That is, $A \subset S$. If a subset A contains only one element (one outcome) then the event is called an elementary event. Elementary events are single elements of the outcome set S.

Ex. 1.16.1. *A geometrical representation of the outcome set of rolling a die twice is given in Fig. 1·1.*

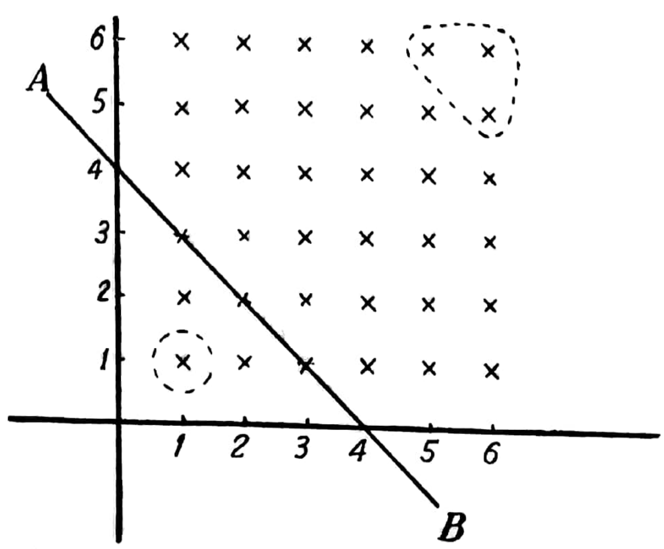


Fig. 1·1.

The event of getting a total 2 (sum of the face numbers is 2) is given by the point encircled in Fig. 1·1. The event of getting a total of 2 or 3 or 4 is given by the set of points below the line AB. The event of rolling 11 or 12 (getting a total of 11 or 12) is given by the set of points inside the dotted closed curve.

Comments. The event of rolling 13 has no point or is a null set. An event which is a null set is called an impossible event. The event of getting a total of either 2 or 3 or 4 or.........or 12 is given by the entire outcome set. This event is sure to happen in any trial. Hence $S \subset S$ is called a sure event.

Ex. 1.16.2. *Consider a dance party consisting of n couples. If the ages of couples are represented as points in a plane then we get n points. Here the outcome set consists of n points in the first quadrant as shown on Fig. 1·2.*

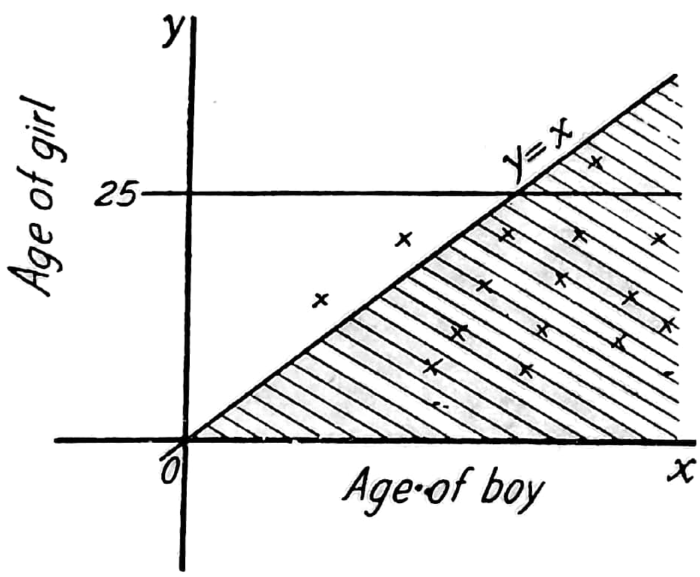


Fig. 1·2.

Consider the event of getting a couple from among the couples where the boy is older than the girl. This event is given by points in the shaded area. The event of getting a couple where the girl's age is less than 25 is given by points below the line $y=25$.

Ex. 1.16.3. *Let the outcome set be symbolically represented by points inside a square S then the events A and B may be represented by the sets of points in the closed curves α and β as shown in Fig. 1·3.*

Such diagrammatic representation of sets is called Venn diagrams. In many problems it will be very convenient to deal with events and probabilities of events if a diagrammatic representation is available.

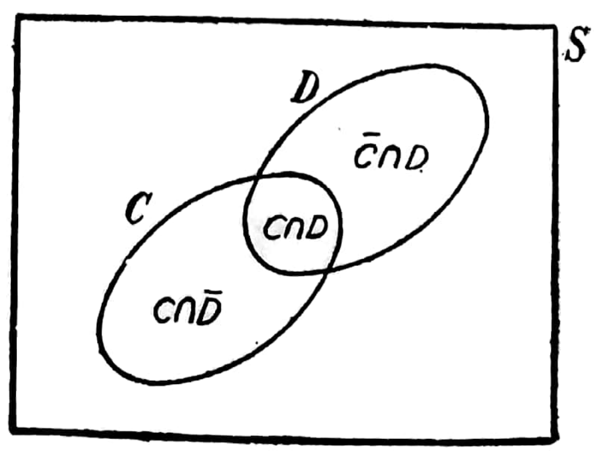


Fig. 1·3

Exercises

1·1. Construct 3 examples each of a set which is *a*) discrete, (*b*) continuous, (*c*) real, (*d*) hypothetical, (*e*) finite, (*f*) infinite.

1·2. Find 5 sets each of which may be considered as consisting of two (*a*) univariate populations, (*b*) bivariate populations.

1·3. In an experiment of rolling a die twice, write down the outcome set as a set of vectors.

1·4. Define the outcome set in an experiment of (*a*) drawing a card from a deck of 52 cards, (*b*) taking two cards successively when the first one is replaced before the second one is drawn, (*c*) taking two cards without replacement.

1·5. An urn contains 10 black balls and 4 white balls. Define the outcome set for the experiment of taking one ball.

1·6. An experiment consists of drawing a card from a well-shuffled deck of 52 cards and then tossing a coin if a red card is drawn. Describe the outcome set or sample space for this experiment.

1·2. VECTORS

1·21. Definition. If the elements of a set are arranged according to some order of succession the set is called a vector. The number of elements in the vector is called the order or the size of the vector. If the elements are put in a column it is called a column vector. In general a vector may be defined as an ordered set. In the following discussion we are interested only in real quantities. We are going to assume that the elements are all real numbers. In general they need not be real numbers.

Ex. 1.21.1. *Consider the integers between 1 and 4 (both inclusive). That is, 1, 2, 3, 4. Then the arrangement (1, 2, 3, 4) is a vector, (1, 3, 2, 4) is another vector, (4, 1, 2, 3) is a third vector etc.*

Comments. From the definition it may be noticed that two vectors are equal if and only if the corresponding elements are all equal. If $V_1=(1, 2, 3)$ and $V_2=(1, 2, 3, 4)$ then equality is not defined since they are of different order. A vector of order one is called a scalar quantity. A reader who is familiar with analytical geometry may notice that vectors may be considered as geometrical points. All the vectors in Ex. 1.21.1 are now vectors of order 4.

Ex. 1.21.2. *Let $x_1\ x_2 \ldots\ldots\ldots, x_n$ denote the yields of wheat at n places. Then the vector $(x_1, \ldots\ldots, x_n)$ denotes a vector of wheat yields.*

Comments. The usual notation for a vector is a simple bracket notation. If the price of every unit of wheat (say per bushel) is Rs. k then the money value of the wheat yields at the different places may be represented by the vector $(kx_1, \ldots\ldots, kx_n)$. In general we can define a scalar multiplication of a vector by the equation

$$k(x_1, \ldots\ldots, x_n)=(kx_1, \ldots\ldots, kx_n).$$

Ex. 1.21.3. *Let $f_1(x), f_2(x), \ldots\ldots\ldots, f_r(x)$ be functions of x then $f=[f_1(x), f_2(x), \ldots\ldots, f_r(x)]$ is a vector of functions, of order r.*

Comments. As the elements of a set can be any well-defined objects the elements of the vector can also be any well-defined objects.

1·22. Addition of Vectors. Let $(a_1, a_2, \ldots\ldots a_n)$ denote the money value of wheat produced at n different places and let $(b_1, b_2, \ldots b_n)$ be the money value of barley produced at the same n places. Then the total money value of wheat and barley produced at the different places is evidently $(a_1+b_1, a_2+b_2, \ldots\ldots\ldots a_n+b_n)$. So in general we will define the sum of two vectors $\mathrm{X}=(x_1, x_2, \ldots \ldots x_n)$ and $\mathrm{Y}=(y_1, y_2, \ldots y_n)$ by $\mathrm{X}+\mathrm{Y}=(x_1+y_1, x_2+y_2, \ldots\ldots x_n+y_n)$.

Ex. 1.22.1. $(1, -1, 0)+(2, 3, 4)=(3, 2, 4)$.

Comments. It may be noticed that only vectors of the same order and type can be added together. If $\mathrm{X}=(x_1, x_2, \ldots\ldots x_n)$ and $\mathbf{0}=(0, 0, 0, \ldots\ldots 0)$ then $\mathrm{X}+\mathbf{0}=\mathrm{X}$. So there exists a vector $\mathbf{0}$ such that $\mathrm{X}+\mathbf{0}=\mathrm{X}$. $\mathbf{0}$ may be called an identity element with respect to the operation addition. A vector with all elements equal to zero is called a null vector and is usually denoted by $\mathbf{0}$.

Ex. 1.22.2. $$\begin{bmatrix} -7 \\ 5 \\ -6 \end{bmatrix} + \begin{bmatrix} 7 \\ -5 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

But $\begin{bmatrix} 1 \\ 5 \\ 7 \end{bmatrix}+[2, 6, -1]$ *is not defined.*

Comments. From this example it may be noticed that for every vector X there exists a vector Y such that $\mathrm{X}+\mathrm{Y}=\mathbf{0}$. This Y may be called an inverse of X with respect to the operation addition.

1·23. Vector Space. Consider a collection of vectors of the same order and category (either all row vectors or all column vectors). Suppose that the sum of any two vectors is also a member of this collection and a scalar multiple of any vector is also a member of this collection. In other words the collection is closed under the operations of addition and scalar multiplication. Then the collection of vectors is called a vector space. That is, if Ω (omega) denotes the collection and if $\mathrm{V}_1 \in \Omega$ (V_1 is an element of Ω) and $\mathrm{V}_2 \in \Omega$ then $\mathrm{V}_1+\mathrm{V}_2 \in \Omega$ and $k\ \mathrm{V}_1 \in \Omega$ where k is any scalar quantity.

Ex. 1.23.1. *Let Ω denote the collection of all possible row vectors of order 5, where the elements are numbers. Evidently Ω is a vector space.*

Comments. A null vector is always an element of a vector space. In this example the vectors in Ω consists of all points in a 5 dimensional Euclidian space. On the other hand if we consider the collection of all vectors of order 5 and all vectors of order 4 then the collection is not a vector space. Sum of any two is not in general an element of this collection because the sum is not in general defined here.

1·24. Independence of Vectors. Consider the vectors $V_1=(1, 2, 3)$ and $V_2=(2, 4, 6)$ then $V_2=2V_1$ or $V_2-2V_1=(0, 0, 0)$ $=\mathbf{0}$. Here V_2 depends on V_1. If for two vectors V_1 and V_2, $V_1+kV_2=\mathbf{0}$ for a non-zero scalar quantity k then V_1 and V_2 are said to be linearly dependent. Let us consider the following vectors,

$$T_1=(1, 2, 3)$$
$$T_2=(1, 2, 4)$$

Let us examine the relation $k_1T_1+k_2T_2=\mathbf{0}$ where k_1 and k_2 are unknown scalar quantities.

$$k_1T_1+k_2T_2=k_1\,(1, 2, 3)+k_2\,(1, 2, 4)$$
$$=(k_1+k_2,\ 2k_1+2k_2,\ 3k_1+3k_2)$$
$$k_1T_1+k_2T_2=\mathbf{0}\Rightarrow k_1+k_2=0,\ 2k_1+2k_2=0,$$
and $$3k_1+4k_2=0\Rightarrow k_1=0=k_2$$

(Here the notation $\Rightarrow$ means 'implies'. For example $2a=2b$ $\Rightarrow a=b$. Similarly $\not\Rightarrow$ means 'does not imply') A number of vectors $V_1, V_2, \ldots\ldots V_r$ each of order n are said to be linearly independent if $k_1V_1+k_2V_2+\ldots\ldots+k_rV_r=\mathbf{0}\Rightarrow k_1=0=k_2=k_3=\ldots\ldots=k_r$ where k_1, $k_2, \ldots\ldots k_r$ are scalar quantities. The number of independent vectors in a vector space is called the dimension or rank of the vector space. If from a set of linearly independent vectors all the vectors in a particular vactor space are available by the operations of addition and scalar multiplication then that set of vectors is called a basis of the vector space generated and the vector space is said to be generated or spanned by the set of basis vectors. The dimension of this vector space is the number of vectors in the generating system or in a basis.

Ex. 1.24.1. *Check whether the vectors* $V_1=(1, 2, 3)$, $V_2=(0, 1, 5)$ *are independent.*

Solution. Let $k_1V_1+k_2V_2=0$ where k_1 and k_2 are scalar quantities.

$$k_1V_1+k_2V_2=k_1\,(1, 2, 3)+k_2\,(0, 1, 5)$$
$$=(k_1,\ 2k_1+k_2,\ 3k_1+5k_2).$$

If $k_1V_1+k_2V_2=(0, 0, 0)=\mathbf{0}$, then

$$\left.\begin{array}{r} k_1=0 \\ 2k_1+k_2=0 \\ 3k_1+5k_2=0 \end{array}\right] \Rightarrow k_1=0 \text{ and } k_2=0$$

The vectors V_1, V_2 are independent.

1·25. Unit Vectors. Consider the following vectors of order n :

$$e_1=(1, 0, 0, \ldots\ldots 0),$$
$$e_2=(0, 1, 0, \ldots\ldots 0),$$
$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$
$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$
$$e_n=(0, 0, \ldots\ldots 1).$$

These vectors $e_1, e_2, \ldots\ldots e_n$ are said to be unit vectors of order n. Any vector

$$X=(x_1, x_2, \ldots\ldots x_n)$$

may be written as a linear combination of these unit vectors in the form,

$$\begin{aligned}
&x_1e_1+x_2e_2+\ldots\ldots+x_ne_n \\
&=x_1(1, 0, 0,\ldots\ldots 0)+x_2(0, 1, 0,\ldots\ldots 0)+\ldots \\
&\qquad\qquad\qquad\qquad\ldots+x_n(0, 0,\ldots\ldots 1) \\
&=(x_1, 0, 0, \ldots\ldots 0)+(0, x_2, 0, \ldots\ldots 0)+\ldots \\
&\qquad\qquad\qquad\qquad\ldots+(0, 0, \ldots\ldots, x_n) \\
&=(x_1, x_2, \ldots\ldots, x_n)=X
\end{aligned}$$

Further it is easily seen that $e_1, e_2, \ldots\ldots e_n$ are independent, and therefore they form a basis of an n-dimensional space.

1·26. Orthogonal Vectors. Two vectors $X=(x_1, x_2, \ldots\ldots x_n)$ and $Y=(y_1, y_2, \ldots\ldots y_n)$ are said to be orthogonal if the sum $x_1y_1=x_2y_2+\ldots\ldots+x_ny_n=0$.

This particular sum is called the inner product of the two vectors X and Y and is usually denoted by XY' (X, Y prime) or YX' (Y, X prime). If in a system of vectors every pair of different vectors are orthogonal then the system is called an orthogonal system of vectors. If Y is the same as X then the inner product XY' reduces to $XX'=x_1^2+x_2^2+\ldots\ldots+x_n^2$.

The positive square root of this XX' is usually called the length of the Vector X and is usually denoted by $\|X\|$ (norm of X).

i.e.,

$$\begin{aligned}
\|X\| &=\sqrt{XX'} \\
&=(x_1^2+x_2^2+\ldots\ldots+x_n^2)^{1/2}.
\end{aligned}$$

It may be noticed that all the vectors $e_1, e_2, \ldots\ldots e_n$ have lengths unity. If in any vector X, $\|X\|=1$ then the vector is called a normal vector. If $Y=\dfrac{X}{\|X\|}$ then Y is a normalized vector of X, since $\|Y\|=1$.

Ex. 1.26.1. *The unit vectors $e_1, e_2, \ldots\ldots e_n$ form an orthogonal system of vectors.*

Comments. If a basis of a vector space is orthogonal it is called an orthogonal basis. It may be noticed that $e_1, e_2, \ldots, e_n$ form an orthonormal basis for an n-dimensional vector space (orthonormal means orthogonal and normal).

Exercises

1·7. A set A is given by $A=\{-1, -5, 0, 6, 7, 8, -10\}$. Find (*a*) all subsets of order 4, (*b*) all vectors of order 4.

1·8. In an experiment of throwing a balanced coin 3 times, find the events of getting (*a*) exactly one head, (*b*) exactly two tails, (*c*) at least two heads. Represent all three as sets of vectors.

1.9. In an experiment of rolling a die once, find the event of getting an even number.

1.10. Find the sum and innerproduct of the vectors $V_1=(1, 2, 5)$, $V_2=(-5, 6, 0)$.

1.11. Construct a vector V_2 orthogonal to the vector $V_1=(5, 6, -7, 0)$, show that every scalar multiple of V_2 is also orthogonal to V_1.

1.12. Find two vectors V_1 and V_2 which are orthogonal and both are orthogonal to the vector $V_3=(1, 1, -1)$.

1.13. Check the independence of the vectors $V_1=(1, 0, -1)$, $V_2=(2, 1, -1)$ and $V_3=(3, 1, -2)$.

1.14. Show that the vectors $V_1=(1, 0, 1)$, $V_2=(2, 0, 1)$ and $V_3=(1, 1, 1)$ can form a basis of a vector space of dimension 3.

1.15. Find the co-ordinates of the vector (1, 1) relative to a basis (1, 0) and (2, 1).

1.3. MATRICES AND LINEAR EQUATIONS

1.31. Linear Equations. Consider the following set of linear equations,

$$y_1=5x_1-2x_2+x_3 \quad \ldots(1)$$

$$y_2=2x_1+7x_2-4x_3 \quad \ldots(2)$$

These equations are completely specified by the coefficients 5, —2, 1 and their order in the first equation and the coefficients 2, 7, —4 and their order in the second equation. That is, the equations are completely specified by the vectors (5, —2, 1) and (2, 7, —4) or if these two vectors are given then a transformation of the quantities x_1, x_2, x_3 into y_1 and y_2 can be written down. If we are interested in the order in which y_1 and y_2 occur, say for example $y_1=5x_1-2x_2+x_3$ is written as the first equation and the other one as the second equation then the equations as well as their order are specified by the arrangement, of vectors in the form,

$$\begin{bmatrix} 5, & -2, & 1 \\ 2, & 7, & -4 \end{bmatrix}$$

where the first row denotes the coefficients in the first equation and the second row denotes the coefficients in the second equation. In general a system of m linear equations in n unknowns and their order are specified by the arrrangement of coefficients in the form,

$$\begin{bmatrix} a_{11}, a_{12}, \ldots, a_{1n} \\ a_{21}, a_{22}, \ldots, a_{2n} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}, a_{m2}, \ldots, a_{mn} \end{bmatrix}$$

where the i^{th} row denotes the coefficients in the i^{th} equation, for $i=1, 2, \ldots, m$ and where for example a_{12} is read as a one, two etc.

1·32. Matrices. The theory of matrices has become a useful tool in almost all branches of applied mathematical activities. Here we will define and discuss some elementary properties of matrices. An arrangement of m row vectors of order n into m rows is called an $m\times n$(m by n) matrix. It may also be considered to be an arrangement of n column vectors of order m into n columns.

Notations. (1) The elements in the vectors are put in brackets, [ ], ( ), etc. For example,

$$A=\begin{bmatrix} 2 & 3 & 5 \\ -1 & 0 & 4 \\ 5 & 7 & -2 \end{bmatrix}$$

is a matrix A which is an arrangement of 3 row vectors of order 3 into 3 rows.

(2) $A=(a_{ij})$. This notation means that the i^{th} now j^{th} column element is a_{ij}. In the above example $a_{11}=2$, $a_{12}=3$, $a_{32}=7$ etc.

(3) $A(m\times n)$. This notation means that there are m rows and n columns in the matrix A.

Comments. When the number of rows=the number of columns=n, the matrix is called a square matrix of order n ; otherwise it is called a rectangular matrix. If all the elements of a matrix are zero it is called a null matrix, and is usually denoted by **0**.

1.33. Samples from a Multivariate Population. From the definition of a multivariate population it is evident that every element in the set designating a k-variate population has k components, namely, the k characteristics which define a k-variate population. If a sample of size n or a subset containing n elements (sometimes called n observations) is taken then we get a matrix of the following form

$$X=\begin{bmatrix} X_{11} & X_{12}\dots X_{1n} \\ X_{21} & X_{22}\dots X_{2n} \\ \vdots & \vdots \quad\quad \vdots \\ X_{k1} & X_{k2}\dots X_{kn} \end{bmatrix}$$

where for example the first column denotes the first element with k components etc.

Ex. 1.33.1. *Consider the height and weight measurements of all students (say n) in a class. This may be written as*

$$X=\begin{bmatrix} x_{11}, & x_{12}, & \dots\dots, & x_{1n} \\ x_{21}, & x_{22}, & \dots\dots, & x_{2n} \end{bmatrix}$$

where the column vectors, $\begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix}, \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix}, \dots\dots, \begin{bmatrix} x_{1n} \\ x_{2n} \end{bmatrix}$ *represent n observations on a bivariate population.*

Comments. If these n students are considered to be a sample of students in a university then the height and weight measurements of the students in this university form a bivariate population. Here every element in the population has two components and hence the population is bivariate.

1·34. Stochastic Matrices. *These are special types of matrices whose elements satisfy some conditions. A singly stochastic matrix $A=(a_{ij})$ may be defined as a matrix whose elements a_{ij} satisfy the conditions. (1) $a_{ij}\geqslant 0$ for all i and j, (2) $\Sigma_{j=1}^{n} a_{ij}=1$ for all i or $\Sigma_{i=1}^{n} a_{ij}=1$ for all j.*

Comments. From conditions (1) and (2) it follows that $0\leqslant a_{ij}\leqslant 1$ for all i and j. Here Σ (sigma) is a notation for a sum. For example,

$$\Sigma_{i=1}^{5} b_i=b_1+b_2+\dots\dots+b_5.$$

$$\Sigma_{i,\ j=1}^{n} b_{ij}=\Sigma_{i=1}^{n}\ \Sigma_{j=1}^{n} b_{ij}=\Sigma_{i=1}^{n}(\Sigma_{j=1}^{n} b_{ij})=\Sigma_{i=1}^{n}(b_{i1}+b_{i2}$$

$$+\dots\dots+b_{in})=\Sigma_{i=1}^{n}\ b_{i1}+\dots\dots+\Sigma_{i=1}^{n}\ b_{in}=(b_{11}+b_{21}+\dots\dots+b_{n1})$$

$$+(b_{12}+\dots\dots+b_{n2})+\dots\dots+(b_{n1}+\dots\dots+b_{nn})=b_{11}+b_{21}+\dots\dots+b_{nn}.$$

$$\Sigma^2_{i=1}\ \Sigma^4_{j=1}\ b_{ij}=b_{11}+b_{12}+b_{13}+b_{14}+b_{21}+b_{22}+b_{23}+b_{24}$$
$$=\Sigma^4_{j=1}\ \Sigma^2_{i=1}\ b_{ij}$$

Ex. 1.34.1.

$$A=\begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

Comments. Here the sum of the elements in any row is unity. If in a stochastic matrix the rows and columns satisfy the conditions that the sum of the elements in any row or column equals unity, then the matrix is called doubly stochastic.

Ex. 1.34.2.

$$B=\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Comments. Here the sum of the elements in any row or column equals unity. Hence B is doubly stochastic. Stochastic matrices are also called Markov matrices. The elements of a Markov matrix satisfy the conditions for probabilities and hence these matrices are called stochastic matrices.

1·35. ALGEBRA OF MATRICES

The development of the ideas in the following sections till the end of this chapter is very fast. If a reader finds it difficult to read the following sections he may omit them and use them as an appendix whenever these ideas are required later. These ideas are used only in a few places in the succeeding chapters.

1.35.1. Addition of Matrices. An operation analogous to the mathematical operation of addition on real numbers will be defined in this section. In order to develop a theory based on these mathematical objects called 'matrices' we have to define mathematical operations on them. We will define the operations 'addition', 'scalar multiplication', 'multiplication' and 'inversion' on matrices.

Definition. Let $A=(a_{ij})$ and $B=(b_{ij})$ then $A+B$ is defined as

$$A+B=(a_{ij}+b_{ij})$$

and it is defined only for matrices of the same category ; that is, $A(m \times n)$ and $B(m \times n)$ may be added, but $A(m \times n)$ and $B(m \times r)$ can not be added up if $n \neq r$ etc.

Ex. 1.35.1. *Let* $A=\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B=\begin{bmatrix} -1 & 0 \\ 2 & 5 \end{bmatrix}$

Then $$A+B=\begin{bmatrix} 1-1 & 2+0 \\ 3+2 & 4+5 \end{bmatrix}=\begin{bmatrix} 0 & 2 \\ 5 & 9 \end{bmatrix}$$

Comments. If O is a null matrix (a matrix with all elements zero) then $A+O=A$. For any matrix A there exists a null matrix with the same number of rows and columns as A such that $A+O=A$. The definition of addition of matrices may be extended as $A+B+C=(a_{ij}+b_{ij}+c_{ij})$ etc. It may be noticed that $A+(B+C)=(A+B)+C$ etc.

1·36. Scalar Multiplication

Definition. $kA=(ka_{ij})$ where $A=(a_{ij})$ and k is a scalar quantity.

Ex. 1.36.1. *Let* $A=\begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & 2 \\ 0 & -1 & 0 \end{bmatrix}$ *and* $k=5$, *then*

$$5A=\begin{bmatrix} 5 & 0 & -5 \\ 10 & 15 & 10 \\ 0 & -5 & 0 \end{bmatrix}$$

Comments. If $k=-1$ then $kA=-A$ and $A+kA=\mathbf{0}$.

or every matrix A there exists an inverse with respect to the operation addition, namely the inverse of A with respect to addition is $(-1)A=-A$.

1·37. The Transpose of a Matrix. Let $A=(a_{ij})$ and A' denote transpose of A then A' is defined as $A'=(a_{ji})$—the ij^{th} element of A' is the ji^{th} element of A.

Ex. 1.37.1. (*a*) $A=\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, $A'=\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$;

(*b*) $A=\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 1 \end{bmatrix}$, $A'=\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 1 \end{bmatrix}$

Comments. The rows of A are the columns of A′ and *vice versa.* In (*b*) A and A′ are the same. In such a case A is called a symmetric matrix. It may be noticed that (A′)′=A always for any matrix A. A vector is a special case of a matrix, so if X is a row vector X′ is a column vector and *vice versa.* A scalar quantity is a matrix having only one row and one column.

1.38. Multiplication of Matrices. The product AB of matrices A and B is defined only if the number of columns of A equals the number of rows of B, that is if A and B are of the form $m \times n$ and $n \times r$ respectively.

Definition. Let $A=(a_{ij})$ and $B=(b_{ij})$ then

$$AB=\left[\sum_{k=1}^{n} a_{ik}\, b_{kj} \right]$$

i.e., the i^{th} row j^{th} column element in AB is the innerproduct of the i^{th} row vector of A and the j^{th} column vector of B.

Ex. 1.38.1. *Let* $A=\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ and $B=\begin{bmatrix} 0 & 1 & 0 \\ 2 & 1 & 5 \\ -2 & 0 & 4 \end{bmatrix}$

then

$$AB=\begin{bmatrix} 1\times 0+2\times 2+3\times(-2), & 1\times 1+2\times 1+3\times 0, & 1\times 0+2\times 5+3\times 4 \\ 4\times 0+5\times 2+6\times(-2), & 4\times 1+5\times 1+6\times 0, & 4\times 0+5\times 5+6\times 4 \end{bmatrix}$$

$$=\begin{bmatrix} -2 & 3 & 22 \\ -2 & 9 & 49 \end{bmatrix}$$

Comments. If A is of the form $m \times n$ (m rows and n columns) and B is of the form $n \times r$ then AB is of the form $m \times r$. In Ex. 1.38.1. A(2×3). B(3×3)=AB(2×3). If A and B are any two matrices then AB need not be equal to BA. If AB=BA then A and B are said to be commutative. If A is a square matrix of order n then A^2 may be defined as A.A and in general A^k=A...A (k times) where k is a positive integer. If A is of the form $m \times n$ and B is of the form $n \times 1$ then AB defines the multiplication of a matrix by a vector.

Ex. 1.38.2. *Let* $A=\begin{bmatrix} 2 & 0 & 5 \\ 1 & 4 & 1 \end{bmatrix}$ and $X=(x_1, x_2, x_3)$ then

$$XA'=(x_1, x_2, x_3)\begin{bmatrix} 2 & 1 \\ 0 & 4 \\ 5 & 1 \end{bmatrix}=(2x_1+5x_3,\ x_1+4x_2+x_3 \quad \text{and}$$

$$AX'=\begin{bmatrix} 2 & 0 & 5 \\ 1 & 4 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}=\begin{bmatrix} 2x_1+5x_3 \\ x_1+4x_2+x_3 \end{bmatrix}$$

Comments. A scalar quantity is a special case of a matrix. So scalar multiplication may be considered to be a special case of matrix multiplication. For example,

$$5A=5\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} A=\begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} A$$

where $$A=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Further inner product XY′ or YX′ of two vectors $X=(x_1, x_2,\ldots,x_n)$ and $Y=(y_1, \ldots, y_n)$ may be considered to be a special case of matrix multiplication.

Ex. 1.38.3. *The following system of linear equations,*

$$a_{11}\, x_1+a_{12}\, x_2+\ldots+a_{1n}\, x_n=y_1$$
$$a_{21}\, x_1+a_{22}\, x_2+\ldots+a_{2n}\, x_n=y_2$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_{m1}\, x_1+a_{m2}\, x_2+\ldots+a_{mn}\, x_n=y_m$$

may be written as a single matrix equation $AX'=Y'$, *where* $A=(a_{ij})$ $X=(x_1, x_2,\ldots x_n)$ and $Y=(y_1, y_2,\cdots y_n)$.

Comments. If y is a null vector then the system of linear equations is called a homogeneous system of linear equations; otherwise it is called a non-homogeneous system. If all the row vectors of A, namely, $(a_{i1}, a_{i2}, \ldots, a_{in})$ for $i=1, 2, 3, \ldots, m$ are independent then the system of equations forms a system of m independent equations. A system of n independent equations in n unknowns has a unique solution, otherwise a solution may not exist and if a solution exists it need not be unique. AX′=Y′ may be considered to be a transformation of $X=(x_1, x_2,\ldots, x_n)$ into $Y=(y_1, y_2, \ldots, y_m)$ where A is the matrix of transformation. AX′=Y′ may also be considered to be a representation of x's and y's where A may be called the matrix in the representation. The above representation is called a simple linear representation.

1.39. Rank of Matrix. The number of independent rows or columns in a matrix A is called the rank of the matrix A. It is the dimension of the vector space generated by the row vectors or the column vectors of A. It can be proved that in any matrix the number of independent rows equals the number of independent columns.

Ex. 1.39.1. *Let* $A=\begin{bmatrix} 1 & 2 & 4 \\ 2 & 4 & 6 \end{bmatrix}$ *then rank of* $A=\rho(A)=1$, *where* $\rho(A)$ *(rho A) is only a notation.*

Comments. The second vector (2, 4, 6) is dependent on the first vector (1, 2, 3) since (2, 4, 6)=2(1, 2, 3).

So the rank is one. The second and the third column vectors are dependent on the first column vector. Hence the number of independent column vectors is also one. The dimension of the vector space generated by the row vectors of order 3=the dimension of the vector space generated by the column vectors of order 2=1. It can be shown that in any matrix, row rank (=the number of independent row vectors)=column rank (=the number of independent column vectors)=the rank of the matrix. The maximum number of independent row or column vectors possible for a matrix A$(m \times n)$ is m if $m \leqslant n$ or n if $n \leqslant m$.

Exercises

1·16. Write down the following system of linear equations in a single matrix equation and also find the rank of the coefficient matrix.

$$\begin{aligned} x_1+\ x_2+3x_3+4x_4&=0 \\ 2x_1+\ x_2 \qquad\ +x_4&=2 \\ -3x_2+5x_3-\ x_4&=5. \end{aligned}$$

1·17. Give two examples each of

(*a*) a singly stochastic matrix,

(*b*) a doubly stochastic matrix.

1·18. If A, B, C are three square matrices show that

(*a*) A+(B+C)=(A+B)+C,

(*b*) A (B+C)=AB+AC,

(*c*) A (BC)=(AB) C.

1·19. In an experiment of throwing a balanced coin thrice, write down the event of getting atleast one head in matrix notation.

1·20. If A and B are two square matrices of order n show that

(*a*) ρ (A+B)$\leqslant\rho$ (A)+ρ (B),

(*b*) ρ (AB)$\leqslant$min. [ρ (A),ρ (B)] where ρ denotes the rank.

1·4. SINGULAR AND NON-SINGULAR MATRICES

The idea of singularity of matrices is closely associated to the linear independence of vectors. A brief introduction to the singularity of matrices is given here.

1·41. Singular matrices. A square matrix of order n is said to be a singular matrix if the rank of A is less than n. If the rank of A equals n then it is said to be non-singular. If the matrix of a linear transformation is singular the transformation is a singular transformation. If the coefficient matrix of a system of linear equations is singular then the system is said to be a singular system of equations; otherwise it is called a non-singular system.

Ex. 1.41.1. *Let* $A=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix}$ then A is singular.

Comments. Here the third row=twice the first row+3 times the second row, and the first two rows are evidently independent (unit vectors). Hence the rank of $A=\rho(A)=2$.

1·42. Identity Matrix. If in a square matrix of order n, the i^{th} row is the i^{th} unit vector for $i=1, 2, \ldots\ldots n$ then the square matrix is called an identity matrix of order n and is usually denoted by I or I_n.

Ex. 1.42.1. *Let* $I=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ *then I is an identity matrix of order 3.*

Comments. An identity matrix is said to have unities along the leading diagonal. The $(ii)^{th}$ elements in a matrix, for $i=1, 2, \ldots\ldots n$ are said to be the leading diagonal elements. If $A=(a_{ij})$ then $a_{11}, a_{22}, a_{33}, \ldots\ldots a_{nn}$ are the leading diagonal elements of A. If in a matrix A all the non-diagonal elements are zero then the matrix is called a diagonal matrix. Usual notations are D, diagonals $(d_1, d_2, \ldots\ldots, d_n)$ etc. Here $d_1, d_2, \ldots\ldots, d_n$ denotes the diagonal elements (some of the d's may be zero). If all the d's are zero then the matrix is evidently a null matrix and if all the diagonal elements are unities then the matrix is an identity matrix. If all the elements above or below the leading diagonal are zeros then such a matrix is called a triangular matrix. It may be noticed that $IA=AI=A$ where A is a square matrix of order n. Also $D_1D_2=D_3$ where D_1, D_2 and D_3 are diagonal matrices, and $\Delta_1\Delta_2=\Delta_3$ where Δ_1, Δ_2 and Δ_3 (delta three) are triangular matrices of the same type. If A and B are two matrices such that $AB=BA$ then A and B are said to commute. It may be noticed that I_n commutes with any square matrix of order n where I_n is the identity matrix of order n. It can be proved that for any square matrix A there exists two non-singular matrices P and Q such that $PAQ=D$ where D is a diagonal matrix. In such a case D is called the canonical form of A.

Exercises

1·21. If A and B are two non-singular matrices and if AC, CB are defined then show that $\rho(AC)=\rho(CB)=\rho(C)$.

1·22. If A and B are square matrices of order n such that $AB=\mathbf{0}$ then show that $A=\mathbf{0}$ or $B=\mathbf{0}$ or A and B are both singular.

1·23. Show that $\rho(AA')=\rho(A'A)=\rho(A)$.

1·24. If $A=\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and if $AB=I$ find out B.

1·5. GEOMETRY OF VECTORS AND MATRICES

1.50. Introduction. So far we have been considering some elementary properties of vectors and matrices. In this section we will consider the geometrical representation of vectors. The reader may be familiar with the geometrical representation of vectors of order one, that is, scalar quantities. In this section, for convenience, we are going to assume that the elements are all real numbers. Vectors of order one, for example, 5, 7, -2 etc. may be represented as points in a one demensional space (or a line).

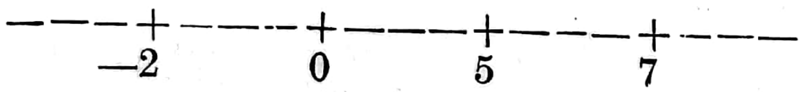


Fig. 1·4.

If we consider vectors or order 2, for example $(-1, 2)$, $(3, 5)$ etc, these may be represented as points in a two dimensional space.

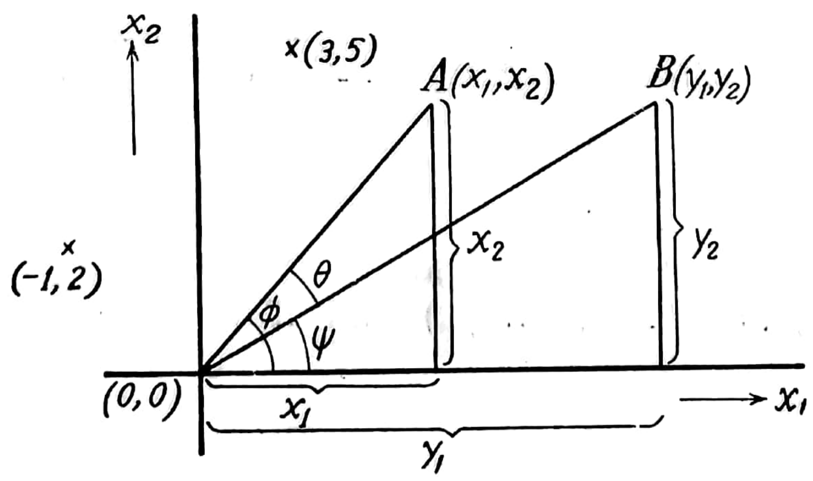


Fig. 1.5.

Here the first element in the vector (the first co-ordinate) is taken along one axis (say the X_1-axis) and the second co-ordinate is taken along the other axis (say the X_2-axis). For example in the vector (3, 5), 3 is the X_1 co-ordinate and 5 is the X_2 co-ordinate. Any vector (x_1, x_2) of order 2 may be considered to be a point in a two dimensional space or a plane. In general a vector

$(x_1,\ldots,x_n)$ of order n may be considered to be a point in an n-dimensional space. (In the geometrical language 'dimension' refers to only the order of the vectors whereas in algebraic language 'dimension' of a vector space means the number of linearly independent vectors which generate the vector space under consideration. The reader may notice the meaning of dimension in the two usages).

1.51. Length of the Vector. The distance between the point represented by a vector and the origin may be taken as the length of a vector. For example if $X=(x_1, x_2)$ is a vector of order 2, that is, a point in a two dimensional space as shown in Fig. 1·5 then the square of the distance between the point A and the origin (*i.e.*, $OA^2)=x_1^2+x_2^2$.

$\| X \|^2=x_1^2+x_2^2=XX'=$square of the length of the vector X. In general if $X=(x_1,\ldots,x_n)$ is a vector of order n then its length may be defined as,

$$\| X \|^2=x_1^2+x_2^2+\ldots+x_n^2=XX'.$$

Evidently when $X=(0, 0,\ldots,0)$, *i.e.*, a null vector then $\| X \|=0$; otherwise $\| X \|>0$.

1.52. Innerproduct. The innerproduct of two vectors $X=(x_1, x_2)$ and $Y=(y_1, y_2)$ was defined as $XY'=x_1y_1+x_2y_2$. In Fig. 1.5,

$$\cos\theta=\cos(\phi-\psi)=\cos\phi\ \text{coe}\ \psi+\sin\phi\sin\psi$$

$$=\frac{x_1\,y_1}{\sqrt{(x_1^2+x_2^2)}\ \sqrt{(y_1^2+y_2^2)}}+\frac{x_2\,y_2}{\sqrt{(x_1^2+x_2^2)}\sqrt{(y_1^2+y_2^2)}}$$

$$=\frac{XY'}{\| X \|\,.\,\| Y \|}$$

where θ(theta), ϕ(phi), ψ(psi) are all Greek letters. In general if $X=(x_1,\ldots, x_n)$, and $Y=(y_1,\ldots,y_n)$ and if θ is the angle between them, then

$$\cos\theta=\frac{XY'}{\| X \|\,.\,\| Y \|}.$$

But $-1 \leqslant \cos\theta \leqslant 1$ and therefore $XY' \leqslant \| X \|\,.\,\| Y \|$ (This inequality is also known as Schwartz's inequality); when $XY'=0$, *i.e.*, when the vectors are orthogonal $\theta=90°$.

1.53. Addition and Scalar Multiplication. It may be easily noticed that if $X=(x_1, x_2)$ and $Y=(y_1, y_2)$ are as shown in Fig. 1·6 then $X+Y=(x_1+y_1, x_2+y_2)$ is a vertex of the parallelo-

gram as shown in Fig. 1·6, and hence $\| X+Y \|$ is the length of the diagonal OC. From the property of a triangle,

$$\| X+Y \| \leqslant \| X \| + \| Y \| .$$

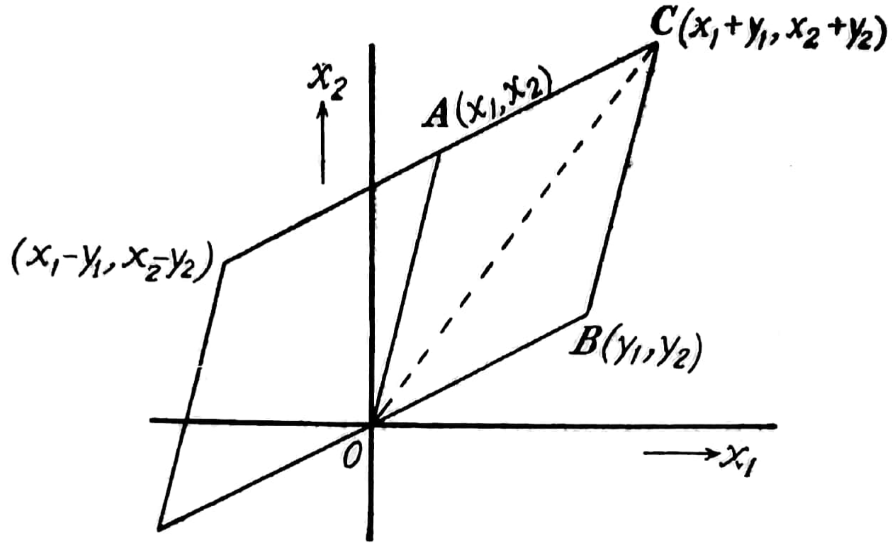


Fig. 1.6.

It is easily seen from Fig. 1·7 that $kX=k(x_1, x_2)$ is a point A_1 on the line OA where A is the point $X=(x_1, x_2)$ and k is a scalar quantity.

$$\| kX \| = \| k \| . \| X \| .$$

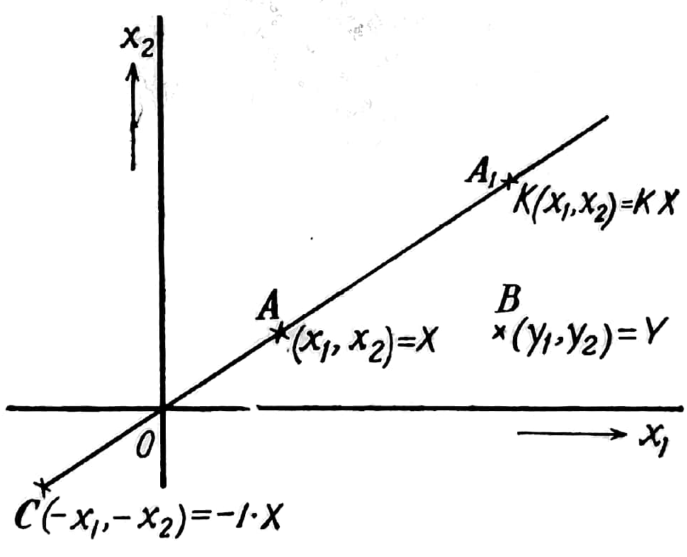


Fig. 1·7.

If X and Y are two points as in Fig. 1.7 then any point in the 2-dimensional space may be obtained as a linear combination of X and Y, *i.e.*, two independent vectors (vectors which are not points on the same line) of order two generate the entire two dimensional space in the sense that any point in the space may be obtained as a linear combination of X and Y or any point Z may be written as

$$Z=aX+bY$$

where a and b are scalar quantities. All these results may be generalized to an n-dimensional space, *i.e.*, n independent vectors of order n generate the entire n-dimensional space.

1·54. Some useful results. It is seen that the length $\| X \|$ of a vector X satisfies the following conditions:

(1) $\| X \| > 0$ and $\| X \| = 0 \Leftrightarrow X = \mathbf{0}$.

($\Leftrightarrow$ means, 'implies both ways' or $\| X \| = 0$ if and only if X is a null vector).

(2) $\| kX \| = | k | . \| X \|$ where k is a scalar quantity.

(3) $\| X+Y \| \leqslant \| X \| + \| Y \|$ where Y is another vector of the same order. Any function of the elements of X satisfying these conditions is called a 'norm' of the vector $X=(x_1, x_2, \ldots\ldots, x_n)$ and is usually denoted by $\| X \|$. Evidently the length of X is also a norm. Other examples for 'norm' of X are

(a) $max_i \; | x_i | = \| X \|$ say

where $max_i \; | x_i |$ means the largest of the magnitudes of $x_1, x_2, \ldots x_n$.

For example if $X=(-1, 2, -5)$ then $max_i \; | x_i | = 5$.

(b) $\Sigma_i \; | x_i | = \| X \|_1$ say.

(c) $\{\Sigma_i \; | x_i |^p\}^{\frac{1}{p}} = \| X \|_p$ say ($p \geqslant 1$).

i.e., when $p=2$, $\| X \|_p^2$ gives the square of the length of X, namely, $\Sigma_i \; | x_i^2 | = x_1^2 + x_2^2 + \ldots\ldots + x_n^2$ (Here all x's are assumed to be real).

In a similar way 'norm' of a square matrix A, say $\| A \|$, may be defined as a function of the elements a_{ij}'s of A, satisfying the conditions.

1. $\| A \| > 0$ and $\| A \| = 0 \Leftrightarrow A = \mathbf{0}$.
2. $\| kA \| = | k | . \| A \|$ where k is a scalar quantity.
3. $\| A+B \| \leqslant \| A \| + \| B \|$ where B is another square matrix of the same order as A.
4. $\| AB \| \leqslant \| A \| . \| B \|$.

The following are some examples for $\| A \|$

(α) $max_i \sum_{j=1}^{n} | a_{ij} | = \| A \|_{I}$ say.

(β) $max_j \sum_{i=1}^{n} | a_{ij} | = \| A \|_{II}$ say.

(γ) $\sqrt{\sum_{ij} |a_{ij}|^2} = \| A \|_{III}$ say.

(δ) $n \; max_{ij} |a_{ij}| = \| A \|_{IV}$ say.

1·55. Linear Transformations

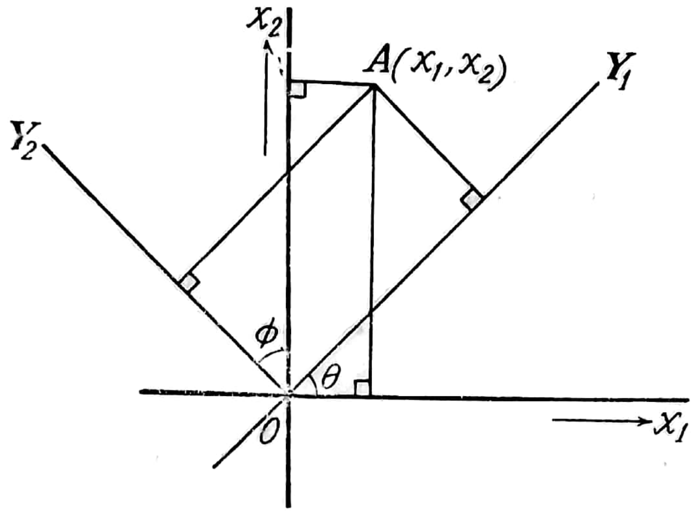


Fig. 1·8.

Let $A=(x_1, x_2)$ be a point in a two dimensional space with reference to the rectangular axes of co-ordinates OX_1 and OX_2 as shown in Fig. 1·8. Let OY_1 and OY_2 be two lines passing through O making angles θ with OX_1 and ϕ with OX_2 as shown in Fig. 1·8. The same point (x_1, x_2) may be written as

(y_1, y_2) with reference to the axes of co-ordinates OY_1 and OY_2. It may be easily seen that

$$Y_1 = a_{11}x_1 + a_{12}x_2$$

and
$$Y_2 = a_{21}x_1 + a_{22}x_2$$

where a_{11}, a_{12}, a_{21}, a_{22} are functions of θ and ϕ

i.e., $$AX' = Y \text{ where } A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, X = (x_1, x_2) \text{ and } Y = (y_1, y_2)$$

in general represents only a change in the co-ordinate system. It may be further noticed that when $\theta = \phi$ or when the angle between OY_1 and OY_2 is $\pi/2 = 90°$ then the matrix A satisfies the condition.

$$AA' = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

In this case the transformation is called an orthogonal transformation. An orthogonal transformation is nothing but a rotation of the axes of coordinates. In general a linear transformation $AX' = Y'$ where A is a non-singular matrix of order n and X and Y are vectors of order n represents a change in the co-ordinate system. If the transformation is orthogonal or if $AA' = 1$ then the transformation is a rotation of the co-ordinate system.

Exercises

1.25. If the following norms $\| A \|_{\mathrm{I}} = \max_i \sum_{k=1}^{n} | a_{ik} |$, $\| A \|_{\mathrm{II}} = \max_k \sum_i | a_{ik} |$, $M(A) = n \max_{ij} | a_{ij} |$ and $N(A) = (\sum_{ij} | a_{ij} |^2)^{1/2}$ are given show that (a) $(1/n) M(A) \leqslant \| A \|_{\mathrm{I}} \leqslant M(A)$; (b) $(1/n) M(A) \leqslant \| A \|_{\mathrm{II}} \leqslant M(A)$; (c) $(1/n)^{1/2} N(A) \leqslant \| A \|_{\mathrm{I}} \leqslant (n)^{1/2} N(A)$; (d) $(1/n)^{1/2} N(A) \leqslant \| A \|_{\mathrm{II}} \leqslant (n)^{1/2} N(A)$.

1.26. An elementary matrix is defined as a matrix obtained by (1) multiplying any row (column) by a scalar quantity, (2) interchanging any two rows (columns), (3) adding a linear combination of rows (columns) to a row (column) of an identity matrix. Multiplication of a matrix by an elementary matrix is called an elementary transformation. Show that any square matrix A may be reduced to a diagonal matrix D by elementary transformations.

1.6. DETERMINANTS

1.60. Introduction. A square matrix is seen to be an arrangement of n^2 quantities into n rows and n columns. In this section we will define a function of the elements of a square matrix. In the previous section we have seen a class of functions defined by their general properties or by a set of axioms or assumptions. Here also the determinant of a square matrix A will be defined as a function of the row or column vectors of A satisfying a set of axioms.

1.61. Some Axioms. Let $\alpha_1, \alpha_2, \ldots\ldots\ldots, \alpha_n$ be the row (or column) vectors of a square matrix A then the determinant of A (denoted by det A or $|A|$) is defined as a function of $\alpha_1, \alpha_2, \ldots\ldots \alpha_n$ satisfying the following conditions.

1. $\det(\alpha_1, \alpha_2, \ldots\ldots, c\alpha_i, \ldots\ldots\ldots \alpha_n) = c.\det(\alpha_1, \alpha_2 \ldots \alpha_i \ldots \alpha_n)$ where c is a scalar quantity.

2. $\det(\alpha_1, \alpha_2 \ldots\ldots, \alpha_i + \alpha_j, \ldots \alpha_n) = \det(\alpha_1, \alpha_2 \ldots \alpha_i \ldots \alpha_j \ldots \alpha_n)$

3. $\det(\alpha_1, \alpha_2, \ldots\ldots \beta_i + \gamma_i, \ldots\ldots \alpha_n) = \det(\alpha_1, \alpha_2, \ldots\ldots \beta_i \ldots\ldots \alpha_n) + \det(\alpha_1, \alpha_2, \ldots\ldots \gamma_i \ldots\ldots \alpha_n)$

4. $\det(e_1, e_2 \ldots\ldots\ldots\ldots\ldots\ldots e_n) = 1.$

Axiom (1) indicates that if any row (or column) of A is multiplied by a scalar quantity c it is equivalent to multiplying the determinant by the same scalar quantity. Axiom (2) says that the value of a determinant is not altered if any row (column) is added to another row (column). The third axiom implies that if any row (column) is split up into two vectors, β_i and γ_i such that $\alpha_i = \beta_i + \gamma_i$ then the determinant can be considered to be the sum of two determinants, one with α_i replaced by β_i and the other with α_i replaced by γ_i. Condition (4) is equivalent to saying

that if the various rows (columns) are the various unit vectors in the natural order then the value of the determinant is unity. Some of the following theorems will enable us to evaluate the determinant of a square matrix A.

Theorem 1.6.1. If one row (column) of a square matrix A is null then $|A|=0$. This follows immediately from axiom (1) by taking $c=0$.

Theorem 1.6.2. The determinant of a singular matrix is zero.

Proof. A singular matrix, according to the definition has dependent rows. So at least one of the rows may be written as linear combinations of others. Applying axioms (2) and (1) this row can be reduced to a null vector. Hence the result.

Theorem 1.6.3. The value of a determinant is not changed if a constant multiple of one row (column) is added to another row (column). This can be easily shown by applying axioms (2) and (1).

Theorem 1.6.4. If two rows (columns) of a determinant are interchanged then the magnitude of the determinant is not changed but the sign is changed.

Proof. $|A|=D(\alpha_1, \alpha_2, \ldots \alpha_i, \ldots \alpha_j, \ldots \alpha_n)$ say

$$
\begin{aligned}
D(\alpha_1, \alpha_2, \ldots \alpha_i, \ldots \alpha_j, \ldots \alpha_n) &= D(\alpha_1, \ldots \alpha_i, \ \ldots \alpha_i+\alpha_j \ldots \alpha_n) && \text{axiom 2} \\
&= -D(\alpha_1, \ldots \alpha_i, \ldots -\alpha_i-\alpha_j, \ldots \alpha_n) && \text{axiom 1} \\
&= -D(\alpha_1, \ldots -\alpha_j, \ldots -\alpha_i-\alpha_j, \ldots \alpha_n) && \text{axiom 2} \\
&= D(\alpha_1, \ldots \alpha_j, \ldots -\alpha_i-\alpha_j, \ldots \alpha_n) && \text{axiom 1} \\
&= D(\alpha_1, \ldots \alpha_j, \ldots -\alpha_i, \ldots \alpha_n) && \text{axiom 2} \\
&= -D(\alpha_1, \ldots \alpha_j, \ldots \alpha_i, \ldots \alpha_n) && \text{axiom 1}
\end{aligned}
$$

1.62. Cofactor of a determinant. If the i^{th} row of $|A|$ is replaced by the j^{th} unit vector e_j then the resulting determinant is called the cofactor of a_{ij} and is usually written as $|A_{ij}|$.

Ex. 1.62.1.
$$
\begin{vmatrix} 0 & 1 & 0 \ldots\ldots 0 \\ a_{21} & a_{22} \ldots\ldots\ldots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} \ldots\ldots\ldots & a_{nn} \end{vmatrix} = |A_{12}| = \textit{Cofactor of } a_{12} \textit{ in } |A|.
$$

Theorem 1.6.5. A determinant $|A|$ may be expanded in terms of the cofactors of the elements of any row (column).

Let $A=(a_{ij})$

Let $\alpha_1=(a_{11}, a_{12}, \dots a_{1n})$ (first row vector) $=a_{11}e_1+a_{12}e_2+\dots+a_{1n}e_n$ where $e_1, e_2\dots, e_n$ are the unit vectors.

By applying axiom (3) successively

$$|A|=D(\alpha_1, \alpha_2, \dots\dots \alpha_n)=D(a_{11}e_1, \alpha_2, \dots \alpha_n)+D(a_{12}e_2, \alpha_2, \dots \alpha_n)+\dots+D(a_{1n}e_n, \alpha_2\dots\alpha_n)$$

$$=a_{11}\, D(e_1, \alpha_2, \dots \alpha_n)+a_{12}D(e_2, \alpha_2, \dots \alpha_n)+\dots+a_{1n}\, D(e_n, \alpha_2, \dots \alpha_n)$$

Theorem 1.6.6.

$$a_{i1}\,|A_{k1}|+a_{i2}\,|A_{k2}|+\dots+a_{in}\,|A_{kn}|=0 \text{ if } k\neq i$$

where $A=(a_{ij})$; *i.e.*, the sum of products of the elements of one row (column) and the cofactors of the elements of any other row (column) is zero.

Proof. $a_{i1}\,|A_{k1}|+a_{i2}\,|A_{k2}|+\dots+a_{in}\,|A_{kn}|$

$$=a_{i1}\, D(\alpha_1, \alpha_2\dots\alpha_i, \dots e_1, \alpha_{k+1}, \dots\alpha_n)+a_{i2}\, D(\alpha_1, \dots\alpha_i, \dots e_2, \alpha_{k+1}, \dots\alpha_n)+\dots+a_{in}\, D(\alpha_1, \alpha_2, \dots\alpha_i, \dots e_n, \alpha_{k+1}\dots\alpha_n)$$

$$=D(\alpha_1\, \alpha_2, \dots\alpha_i, \dots a_{i1}\, e_1, \alpha_{k+1}, \dots\alpha_n)+D(\alpha_1, \dots\alpha_i, \dots \alpha_{12}e_2, \alpha_{k+1}\dots\alpha_n)+\dots+D(\alpha_1, \alpha_2, \dots\alpha_i, \dots a_{in}e_n, \alpha_{k+1}, \dots\alpha_n)$$

$$=D(\alpha_1\dots\alpha_i, \dots\alpha_i, \alpha_{k+1}, \dots\alpha_n)=0$$ since the i^{th} and k^{th} rows are the same.

Theorem 1.6.7. $|A|=\sum\limits_{i, j, \dots k}\pm a_{1i}\, a_{2j}\dots a_{nk}$ $i\neq j\neq\dots\neq k$ (all different)

Proof. Let $\alpha_1=a_{11}e_1+a_{12}e_2+\dots+a_{1n}, e_n$.

$$|A|=a_{11}\,|A_{11}|+a_{12}\,|A_{12}|+\dots+a_{1n}\,|A_{1n}| \quad \dots(1)$$

But $\alpha_i=\alpha_{i1}e_1+a_{i2}e_2+\dots+a_{in}e_n$; $i=1, 2\dots n$.

$$|A|=a_{11}D(e_1, \alpha_2, \dots\alpha_n)+\dots+a_{1n}D(e_n, \alpha_2, \dots\alpha_n).$$

Now expanding $\alpha_2, \dots\alpha_n$ and proceeding in a similar way as in (1), $|A|=\sum\limits_{i, j, \dots k} a_{1i}a_{2j}\dots a_{nk}D(e_i, e_j\dots e_k)$

where the summation for all $i, j, \dots k=1, 2, \dots n$ or each suffix varies from 1, to n.

$D(e_i, e_j, \dots e_k)$ is a determinant where the rows (columns) are $e_1, e_2, \dots, e_n$ in some order.

Therefore $D(e_i, e_j, \dots e_k)=\pm 1$ or 0.

Hence $|A|=\sum\limits_{i, j, \dots k}\pm a_{1i}a_{2j}\dots a_{nk}$; $i\neq j\neq\dots\neq k$ (all different)

i.e., $|A|$ contains a number of terms where every term contains one and only one element from every row and column of A.

From the expansion of $|A|$ it can be easily seen that cofactor of a_{ij} in a matrix $A=(a_{ij})$ is the determinant obtained by deleting the i^{th} row and i^{th} column of A, multiplied by $(-1)^{i+j}$. The determinant obtained from A by deleting the i^{th} row and j^{th} column of A is called the minor of a_{ij} in A.

i.e., Cofactor of $a_{ij}=(-1)^{i+j}$ minor of a_{ij}.

1.63. Simple Rules for Evaluation of Determinants. Some simple rules for evaluation of determinants of order 2 and 3 are suggested here. Determinants of higher orders may be evaluated by using the theorems 1.6.1 to 1.6.7. A determinant of order 2 may be evaluated as

$$|A|=\begin{vmatrix} a & b \\ c & d \end{vmatrix}=ad-bc$$

This may be easily verified by spliting the rows $\alpha_1=(a, b)$ and $\alpha_2=(c, d)$ in terms of unit vectors and expanding $|A|$.

A determinant of order 3 may be evaluated as follows :

$$|A|=\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}=a_1\begin{vmatrix} b_2 & b_3 \\ c_2 & b_3 \end{vmatrix}-a_2\begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix}+a_3\begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}.$$

The determinants of the various submatrices

$$\begin{bmatrix} b_2 & b_3 \\ c_2 & c_3 \end{bmatrix}, \begin{bmatrix} b_1 & b_3 \\ c_1 & c_3 \end{bmatrix}, \begin{bmatrix} b_1 & b_2 \\ c_1 & c_2 \end{bmatrix}$$

are the minors of a_1, a_2 and a_3 respectively.

Equation (3) is the expansion of $|A|$ in terms of the elements in the first row and their cofactors. Similarly $|A|$ may be expanded in terms of the elements and their cofactors of any other row (column).

$$\begin{aligned} |A| &= a_1(b_2c_3-b_3c_2)-a_2(b_1c_3-b_3c_1)+a_3(b_1c_2-b_2c_1). \\ &= a_1b_2c_3+a_2b_3c_1+a_3b_1c_2-a_1b_3c_2-a_2b_1c_3-a_3b_2c_1 \end{aligned}$$

There are six terms in the expansion and further each term contains one and only one element from every row and column of

A. From the nature of the terms in the expansion of $|A|$ a mechanical way of writing down the terms in the expansion of a third order determinant may be suggested as follows :

Let $$|A| = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ & e_2 & c \end{vmatrix}.$$

Augment the columns of A with any two consecutive columns of A. For example if A is augmented with the first and second columns of A then the resulting arrangement is as follows :

$$\begin{array}{ccccc} a_1 & a_2 & a_3 & a_1 & a_2 \\ \\ b_1 & b_2 & b_3 & b_1 & b_2 \\ \\ c_1 & c_2 & c_3 & c_1 & c_2 \end{array}$$

Then multiply and add elements joined by straight lines. From this sum subtract the sum of products of elements joined by dotted lines.

i.e., $|A| = a_1 b_2 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_3 b_2 c_1 - a_1 b_3 c_2 - a_2 b_1 c_3.$

Determinants in general may be evaluated by using the axioms or theorems given in this section.

Ex. 1.63.1. *Evaluate the determinant*

$$|A| = \begin{vmatrix} 1 & 0 & -1 & 2 \\ 2 & -1 & 1 & 0 \\ 3 & 0 & 2 & -4 \\ 1 & 2 & 4 & 5 \end{vmatrix}$$

Solution : Add (-2) times the first row to the second row (the value of $|A|$ is not changed).

$$|A| = \begin{vmatrix} 1 & 0 & -1 & 2 \\ 2 & -1 & 1 & 0 \\ 3 & 0 & 2 & -4 \\ 1 & 2 & 4 & 5 \end{vmatrix} = \begin{vmatrix} 1 & 0 & -1 & 2 \\ 0 & -1 & 3 & -4 \\ 3 & 0 & 2 & -4 \\ 1 & 2 & 4 & 5 \end{vmatrix}.$$

Similarly add (-3) times the first row of A to the third row and (-1) times the first row of A to the 4th row of A, then

$$|A| = \begin{vmatrix} 1 & 0 & -1 & 2 \\ 0 & -1 & 3 & -4 \\ 0 & 0 & 5 & -10 \\ 0 & 2 & 5 & 3 \end{vmatrix}.$$

Expand $|A|$ in terms of the elements of the first column and their cofactors.

$$|A| = \begin{vmatrix} -1 & 3 & -4 \\ 0 & 5 & -10 \\ 2 & 5 & 3 \end{vmatrix} + 0 + 0 + 0$$

$$= \begin{vmatrix} -1 & 3 & -4 \\ 0 & 5 & -10 \\ 2 & 5 & 3 \end{vmatrix}.$$

And 2 times the first row to the third row, then

$$|A| = \begin{vmatrix} -1 & 3 & -4 \\ 0 & 5 & -10 \\ 0 & 11 & -5 \end{vmatrix}.$$

Expand $|A|$ in terms of the elements for the first column and their cofactors.

$$|A| = (-1)\begin{vmatrix} 5 & -10 \\ 11 & -5 \end{vmatrix} + 0 + 0$$

$$= -\begin{vmatrix} 5 & -10 \\ 11 & -5 \end{vmatrix}$$

$$= -(-25 + 110) = -85.$$

Comments. It is easy to see that

(1) $|I| = 1$

(2) $|D| = d_1, d_2 \ldots\ldots d_n$, where D is a diagonal matrix with diagonal elements $d_1, d_2, \ldots, d_n$.

(3) $|T| = t_1, t_2 \ldots\ldots t_n$, where T is a triangular matrix with the diagonal elements $t_1, t_2, \ldots t_n$.

For example :
$$\begin{vmatrix} 2 & 0 & 0 & 0 \\ -1 & 7 & 0 & 0 \\ 0 & 4 & 3 & 0 \\ 5 & -4 & 0 & 4 \end{vmatrix} = 2\times 7\times 3\times 4 = 168.$$

(4) $|AB| = |A|\ |B|$ (*A and B are square matrices*).

(5) $|A'| = |A|$ where A′ is the transpose of A. This follows from symmetry or from the definition itself.

(6) $|P| = \pm 1$ If P is an orthogonal matrix.

Proof. If P is orthogonal $P'P=1$

$$|PP'| = |P|\ |P'|\ |P|^2 = 1. \textit{ Hence the result}$$

Exercises

1.27. Evaluate the determinant of the square matrix A, where

$$A = \begin{vmatrix} 5 & 6 & 3 & 2 \\ 0 & 1 & 0 & 4 \\ 1 & 2 & 3 & 4 \\ 6 & 0 & 4 & 0 \end{vmatrix}$$

1.28. If $B=PAP'$ where P is an orthogonal matrix, show that
$$|B| = |A|.$$

1.29. When a matrix A is reduced to a diagonal matrix D by elementary transformations, D is called the canonical form of A. Show that the canonical form of a singular matrix has at least one zero diagonal element.

1.7. INVERSE OF A MATRIX

The inverse A^{-1} of a non-singular matrix A is defined as the matrix $A^{-1}=(a^{ij})$ where (a^{ij}) is the cofactor of the $(ji)^{th}$ element of A (that is, a_{ji}), divided by the determinant of A.

i.e.,
$$A^{-1}=(a^{ij})=\left(\frac{|A_{ji}|}{|A|}\right).$$

$|A_{ji}|$ is the cofactor of a_{ji} in A.

Ex. 1.7.1. *Let*

$$A = \begin{vmatrix} 1 & 2 & 3 \\ -1 & 2 & 4 \\ 3 & 0 & 5 \end{vmatrix}$$

Then $A^{-1}=\begin{vmatrix} a^{11} & a^{12} & a^{13} \\ a^{21} & a^{22} & a^{23} \\ a^{31} & a^{32} & a^{33} \end{vmatrix}$

where $a^{11}=|A_{11}|/|A|=\begin{vmatrix} 2 & 4 \\ 0 & 5 \end{vmatrix} \div |A|,$

$a^{12}=\begin{vmatrix} 2 & 3 \\ 0 & 5 \end{vmatrix} \div |A|$ etc.

Comments. If A is singular then $|A|=0$. Therefore the inverse is not defined for a singular matrix A.

Theorem 1.7.1. $A.A^{-1}=I=A^{-1}A$.

Proof. Let $A=(a_{ij})$

$$A^{-1}=\frac{1}{|A|}\begin{vmatrix} |A_{11}| & |A_{21}| & \cdots & |A_{n1}| \\ |A_{12}| & |A_{22}| & \cdots & |A_{n2}| \\ \vdots & \vdots & & \vdots \\ |A_{1n}| & |A_{2n}| & \cdots & |A_{nn}| \end{vmatrix}$$

$$A.A^{-1}=\frac{1}{|A|}\begin{vmatrix} a_{11} & a_{12} \cdots a_{1n} \\ a_{21} & a_{22} \cdots a_{2n} \\ \vdots & \vdots \quad \vdots \\ a_{n1} & a_{n2} \cdots a_{nn} \end{vmatrix}\begin{vmatrix} |A_{11}| & |A_{21}| & \cdots & |A_{n1}| \\ \vdots & \vdots & & \vdots \\ |A_{1n}| & |A_{2n}| & \cdots & |A_{nn}| \end{vmatrix}$$

$$=\frac{1}{|A|}\begin{bmatrix} |A| & 0 & 0\ldots0 \\ 0 & |A| & 0\ldots0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots\cdots & |A| \end{bmatrix} \quad \text{(Theorem 1·6·6).}$$

$=I.$

Similarly $A^{-1}A=I$

Comments. It may be easily seen that

(1) If $D=\text{diag}\ (d_1, d_2,\ldots d_n)$, $d_i \neq 0$ for $i=1, 2,\ldots n$ then
$D^{-1}=\text{diag}\ (d_1^{-1}, d_2^{-1},\ldots d_n^{-1})$

(2) $(AB)^{-1}=B^{-1}.A^{-1}$.

(3) If P is orthogonal $P'=P^{-1}$.

(4) $I^{-1}=I$, (5) If $PAP'=D$ then $|A|=|D|$ where P is an orthogonal matrix and D is a diagonal matrix.

1·71. Solutions of Linear Equations. In a system of linear equations $AX'=Y'$ if A is a square matrix of order n and rank n (*i.e.*, the system is a non-singular system of linear equations)

then $X'=A^{-1}Y'$ (obtained by pre-multiplying $AX'=Y'$ by A^{-1}) It can be proved that a system of linear equations $AX'=Y'$ is consistent (or has a solution, where A and Y are known and X is unknown), if the ranks of A and C, where C is a matrix obtained by augmenting A with Y', are equal. If the number of independent equations is less than the number of unknowns, then the equations have an infinite number of solutions. For a system $AX'=\mathbf{0}$ if the rank of A is r then the total number of independent solutions is $n-r$. If $r=n$ then there is only a **trivial** solution $X=\mathbf{0}$. The general solution of a non-homogeneous system of linear equations $AX'=Y'$ where A and Y are known, is obtained by adding a particular solution of $AX'=Y'$ to the general solution of the corresponding homogeneous system $AX'=0$.

Ex. 1.71.1. *Solve* $2x+3y=5$

Sol: Let
$$y=0$$
then
$$x=\frac{5}{2}$$

For various values of y various values of x are obtained.

This system has an infinite number of solutions.

Ex. 1.71.2. *Solve*
$$2x+3y=5$$
$$x+y=2$$

i.e.,
$$\begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

or
$$AX'=Y'.$$

Sol : Here A is non-singular [since A has rank 2, or the vectors (2, 3) and (1, 1) are independent]. Therefore

$$\begin{bmatrix} x \\ y \end{bmatrix}=A^{-1}Y'$$

$$=\begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}^{-1}\begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

$$=\begin{bmatrix} -1 & 3 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

Since $|A|=-1$

$$=\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

or
$$x=1$$
and
$$y=1.$$

The solution is unique.

For a simple system like, two equations in two unknowns, the classical method of elimination of variables is easier than inverting a matrix. But the above method is a general method which can be applied when the system of equations involves a number of equations and a number of unknowns.

1·72. Quadratic and Bilinear Forms. A few definitions and some elementary properties of quadratic and bilinear forms are given in this section.

Definition. A quadratic form XAX′ is said to be positive definite, positive semi-definite, negative definite, negative semi-definite, if $XAX'>0, \geqslant 0, <0, \leqslant 0$ for any real non-null vector X, respectively, where X is a row vector, X′ is its transpose and A is a symmetric matrix. That is, for any non-null real vector X, if,

$XAX'>0$ then XAX′ is positive definite,
$XAX'\geqslant 0$,, positive semi-definite,
$XAX'<0$,, negative definite, and
$XAX'\leqslant 0$,, negative semi-definite.

Without loss of generality A is assumed to be a symmetric matrix.

$XAX'=\Sigma_{i=1}^{n} \; \Sigma_{j=1}^{n} \; a_{ij} \cdot x_i x_j$. Since $x_i x_j = x_j x_i$, if $a_{ij} \neq a_{ji}$ (that is, if the matrix A is not symmetric) a new matrix B can be constructed where $b_{ij}=b_{ji}=(a_{ij}+a_{ji})/2$ and here B is symmetric and further $XAX'=XBX'$. Hence the assumption of symmetry in A is not a restriction. If $X=(x_1, \ldots\ldots, x_n)$ and $Y=(y_1, y_2, \ldots\ldots \ldots y_n)$ are two vectors of order n and if A is a square matrix of order n then XAY′ is called a bilinear form.

Exercises

1·30. Show that if A and B are non-singular square matrices of order n,

(*a*) $(AB)^{-1}=B^{-1}A^{-1}$,

(*b*) $(A^{-1})^{-1}=A$.

1·31. Show that $|A^{-1}|=\dfrac{1}{|A|}$ if $|A| \neq 0$.

1·32. The trace of a square matrix A is defined as the sum of the leading diagonal elements. Show that $tr(AB)=tr(BA)$.

1·33. If P is an orthogonal matrix show that $trPAP'=trA$.

1·34. If $A=A^2$ then A is called an idempotent matrix. Show that for an idempotent matrix A,

(*a*) trA=rank of A;

(*b*) the only non-singular idempotent matrix is I.

1·35. If A is idempotent and if $A+B=I$, show that

(*a*) B is idempotent,

(*b*) $AB=BA=\mathbf{0}$.

1·36. If X is an $n\times 1$ vector with elements $x_1, \ldots\ldots, x_n$ and if A is an $n\times 1$ vector with elements $a_1, \ldots\ldots a_n$ and if $Y=X'A$ then the derivative of Y with respect to X, denoted by $\partial Y/\partial X$ is defined as,

$$\partial Y/\partial X=\begin{bmatrix} \partial Y/\partial x_1 \\ \partial Y/\partial x_3 \\ \cdots\cdots \\ \cdots\cdots \\ \partial Y/\partial x_n \end{bmatrix}$$

then show that

(1) $\partial Y/\partial X=A$,

(2) generalize this to the case where A is a matrix ;

(3) $\partial Y/\partial X=2X'A$ where $Y=X'AX$ and A is a symmetric matrix ;

(4) $\partial Y/\partial A=2XX'-D(XX')$ where $Y=X'AX$ is a quadratic form and $D(XX')$ is a diagonal matrix whose diagonal elements are the diagonal elements of XX' and $\partial Y/\partial A$ is defined as the matrix with (ij)th element $\partial Y/\partial a_{ij}$. (Assume symmetry for A).

1·37. Let $Y=\theta A'+e$ where Y and e are $1\times n$ vectors, θ is a $1\times p$ vector and A' is a $p\times n$ matrix of rank p. Show that the value $\hat{\theta}$ of θ for which $ee'=\Sigma e_i^2$ is minimized with respect to θ, is given by $\hat{\theta}=YA(A'A)^{-1}$.

BIBLIOGRAPHY

Suggested reading in the theory of Matrices and Vectors.

1. D.K. **Faddeev** & V.N. **Faddeeva.** Computational Methods of Linear Algebra. Translation from Russian, W.H. Freeman and Co., San Francisco, 1963.

2. G. **Hadley.** Linear Algebra. Addison Wesley, Inc. Reading, Mass. 1964.

3. D.C. **Murdoch.** Linear Algebra for Under Graduates. John Wiley & Sons, New York, 1957.

4. J.T. **Schwartz.** Introduction to Matrices & Vectors. McGraw-Hill Book Co., New York, 1961.

5. R.M. **Thrall.** Vector Spaces & Matrices. John Wiley & Sons, New York, 1957.

Suggested reading in elementary Statistics.

6. D.A.S. **Fraser.** Statistics, An Introduction. John Wiley, N.Y., 1958.

7. J.E. **Freund.** Mathematical Statistics. Prentice Hall, Inc, N.J. 1964.

8. P.G. **Hoel.** Introduction to Mathematical Statistics. John Wiley, New York, 1962.

CHAPTER 2

PROBABILITY

2·0. Introduction. In day to day life we make statements like, it is very likely that Kumary Latha will win the coming music competition, tomorrow will probably be a sunny day, the chances are almost nil that a man will live for ever, drug X may be more effective than drug Y in curing disease Z etc. In all these statements there is an element of lack of certainty. Statistics and especially the theory of Probability have a vital role in making decisions in the face of lack of certainty. Probability theory is mainly developed to describe and in some sense measure lack of certainties in situations where there is an element of uncertainty. There are three basic problems in the theory, namely,

(1) to describe the situation clearly or to specify a set on which probability statements are made.

(2) to define a numerical measure for a probability statement and

(3) to evaluate numerically the probabilities in specific situations or for particular events.

Even though the palmists, astrologers and fortune-tellers of ancient India might have used a record of past events to predict the future, the recorded evidence of a systematic study of the present day probability theory is that, it developed as a theory of games of chance in the 17th century when some ardent gamblers consulted mathematicians about dividing the stake money in cases of incomplete games.

2·10. Number of Vectors and Subsets from a given Set. Eventhough this section is not very important in defining probability, it is useful in solving some problems of a probabilistic nature. In chapter 1 we were considering sets, subsets, special types of sets, vectors, matrices etc. We did not consider the number of possible subsets and possible vectors from a given set. Consider the word PET. Let us consider all possible three letter words that can be made from the letters of this word. The words are PET, PTE, EPT, ETP, TEP and TPE. There are six different words possible. If the order in which the letters occur is not important then there is only one arrangement, because all the words have the same letters. Let us take the word CUTE and consider all two letter words possible from this word. The words are given

below. There are 12 words. If the order is not important there are only six arrangements.

PET	CUTE	
3 letter words	*2 letter words*	
PET	CU	UT
PTE	UC	TU
EPT	CT	UE
ETP	TC	EU
TEP	CE	TE
TPE	EC	ET
6	12	

The process of forming 3 letter words from the letters P, E, T, may be considered to be a problem of filling three boxes with three different objects. After having filled the first box the second may be filled with one of the remaining two objects. Similarly the third box may be filled up by the remaining object. There are altogether $3 \times 2 \times 1 = 6$ ways. This problem may also be considered to be a problem of determining the ordered sets of distinct elements from a set of distinct elements P, E and T.

2·11. Number of Vectors from a given Set. The total number of possible vectors of order r with distinct elements from a set of n distinct elements may be defined as the number of permutations of n objects taken r at a time; or it may also be defined as the number of ways in which r distinct objects may be selected out of n different objects, taking into consideration the order in which the elements are taken. This number is called the number of permutations of n objects taking r at a time. The procedure of arranging or selecting the objects in some order is called a permutation. The number of permutations is usually denoted by $(n)r$. Other notations are nP_r, $P(n, r)$ etc.

Theorem 2·11·1. $(n)r = n(n-1)(n-2)\ldots(n-r+1)$

Proof. Let the given set be $S = (a_1, a_2, \ldots, a_n)$ where all the a's are distinct. Let $V = (b_1, b_2, \ldots, b_n)$ where $b_i \in S$ for all i and no two of the b's are the same or $b_1, b_2, \ldots b_r$ are some r of the a's from S. The total number of V's may be determined as follows. One of $a_1, a_2, \ldots a_n$ may be taken as b_1. After having selected b_1 there are $(n-1)$, a's left. Out of these one may be selected for b_2. There are $(n-1)$ ways of doing this since there are $(n-1)$ a's left, etc. Therefore the total number of V's

i.e., $$(n)r = n(n-1)(n-2)\ldots(n-r+1)$$

$$= \frac{n(n-1)(n-2)\ldots(n-r+1)(n-r)\ldots 2.1}{(n-r)\ldots 2.1}$$

$=\frac{n\,!}{(n-r)\,!}$ where n ! (n factorial or factorial n

$=1.2.3...n=n(n-1)...2.1.$)

Ex. 2·11·1. *Let $S=\{1, 0, -1, 5\}$.*

Find out the total number of permutations taking 3 at a time.

Solution. Here $n=4$ and $r=3$ and according to the formula

$(n)r=(4)3=4\times3\times2=24=4\,!/1\,!$

The vectors are given below :

$V_1=(1, 0, -1)$	$V_7=(1, 0, 5)$	$V_{13}=(1, -1, 5)$	$V_{19}=(0, -1, 5)$
$V_2=(1, -1, 0)$	$V_8=(1, 5, 0)$	$V_{14}=(1, 5, -1)$	$V_{20}=(0, 5, -1)$
$V_3=(0, 1, -1)$	$V_9=(0, 1, 5)$	$V_{15}=(-1, 1, 5)$	$V_{21}=(-1, 0, 5)$
$V_4=(0, -1, 1)$	$V_{10}=(0, 5, 1)$	$V_{16}=(-1, 5, 1)$	$V_{22}=(-1, 5, 0)$
$V_5=(-1, 1, 0)$	$V_{11}=(5, 1, 0)$	$V_{17}=(5, 1, -1)$	$V_{23}=(5, 0, -1)$
$V_6=(-1, 0, 1)$	$V_{12}=(5, 0, 1)$	$V_{18}=(5, -1, 1)$	$V_{24}=(5, -1, 0)$

Comments. If $r=4$, that is, if we consider all possible vectors of order 4 then it is easily seen to be

$$4\,!=4\times3\times2\times1=24.$$

In general $(n)n=\frac{n\,!}{(n-n)\,!}=\frac{n\,!}{0\,!}=n\,!$ ($0\,!=1$ is assumed)

Ex. 2·11·2. *Find the number of distinct four letter words that can be made using all the letters of the word "doll".*

Solution. If the two l's are distinct like l_1, l_2 then the total number of words is evidently $4\,!=24$. The two l's may be arranged in 2 ! ways. Since $l_1=l_2=l$ the total number of distinct words $=\frac{4\,!}{2\,!}=12$.

Comments. If the elements of a set are not distinct, in some sense $(n)r$ can be defined by a modification seen from this example. The number of distinct vectors can always be determined even if in the original set some elements are repeated.

Ex. 2·11·3. *Find the number of distinct words that can be made using all the letters of the word "Mississippi".*

Solution. If the four i's in the word are denoted by i_1, i_2, i_3 and i_4, the four s's by s_1, s_2, s_3 and s_4 and the $2p$'s by p_1 and p_2 then the total number of words$=11$! The 4 i's can be arranged in 4 ! ways, the four s's in 4 ! ways, the $2p$'s in 2 ! ways and the one m in 1 ! way. Therefore the total number of distinct words is

$$\frac{11\,!}{4\,!\,4\,!\,2\,!\,1\,!}=34650.$$

2·12. The number of Subsets from a given Set. In permutations the order of the elements of a set is taken into consideration. If the order is not considered the number of ways in which a set of two letters can be taken from a set $S=\{a, b, c\}$ of three letters is evidently 3. The different subsets are $\{a, b\}$, $\{b, c\}$, $\{c, a\}$. Here $\{a, b\}$ and $\{b, a\}$ represent the same subset. The number of ways in which subsets of order r can be taken from a set of order n is called the number of combinations of n taken r at a time and is usually denoted by

$$\binom{n}{r},\ {}^nC_r,\ nCr,\ C(n, r) \text{ etc.}$$

Theorem 2·12.1. $\binom{n}{r}=\dfrac{(n)r}{r\,!}=\dfrac{n\,!}{r\,!\,(n-r)\,!}$

Proof. A total number of $r\,!$ vectors or ordered sets may be formed from a set of order r. Therefore the total number of subsets, without considering the order

$$=\frac{(n)r}{r\,!}=\frac{n\,!}{r\,!\,(n-r)\,!}$$

Theorem 2·12·2. (1) $\binom{n}{1}=n$, (2) $\binom{n}{n}=1$, for n a positive integer

Proof. $\binom{n}{r}=\dfrac{n\,!}{r\,!\,(n-r)\,!}$

when $r=1$, $\binom{n}{r}=\dfrac{n\,!}{1\,!\,(n-1)\,!}=\dfrac{n\,!}{(n-1)\,!}$

$$=\frac{n(n-1)\ldots\ldots2\cdot1}{(n-1)\ldots\ldots2\cdot1}=n.$$

when $r=n$, $\binom{n}{r}=\dfrac{n\,!}{n\,!\,(n-n)\,!}=\dfrac{n\,!}{n\,!\,0\,!}=\dfrac{n\,!}{n\,!}=1.$

Theorem 2·12·3. $\binom{n}{r}=\binom{n}{n-r}$ for n a positive integer and $r=0, 1, 2,\ldots\ldots n$.

Proof. If r is replaced by $n-r$, $\binom{n}{r}$ becomes

$$\binom{n}{n-r}=\frac{n\,!}{(n-r)\,!\,(n-n+r)\,!}=\frac{n\,!}{(n-r)\,!\,r\,!}$$

$$=\frac{n\,!}{r\,!\,(n-r)\,!}=\binom{n}{r}.$$

For example

$$\binom{4}{4}=\binom{4}{4-4}=\binom{4}{0}=1.$$

$$\binom{4}{3}=\binom{4}{4-3}=\binom{4}{1}=4$$

$$\binom{4}{2}=6.$$

This shows that there is symmetry among the various combinations. The number of combinations of n taking one at a time is the same as the number of combinations of n taking $(n-1)$ at a time etc.

Theorem 2·12·4. $\binom{n}{r}=\binom{n-1}{r}+\binom{n-1}{r-1}$ for a positive integer n and for $r=0, 1, 2, \ldots\ldots (n-1)$.

Proof. $\binom{n-1}{r}=\dfrac{(n-1)\,!}{r\,!\,(n-1-r)\,!}=\dfrac{(n-1)\,!}{r\;(r-1)\,!\,(n-1-r)\,!}$

$$\binom{n-1}{r-1}=\frac{(n-1)\,!}{(r-1)\,!\,(n-1-r+1)\,!}$$

$$=\frac{(n-1)\,!}{(r-1)\,!\;(n-1-r+1)[(n-1-r)\,!]}$$

$$\therefore \binom{n-1}{r}+\binom{n-1}{r-1}$$

$$=\frac{(n-1)\,!}{(r-1)\,!\,(n-1-r)\,!}\left[\frac{1}{r}+\frac{1}{n-1+1-r}\right]$$

$$=\frac{n.(n-1)\,!}{r(r-1)\,!\,(n-r)(n-1-r)\,!}$$

$$=\frac{n\,!}{r\,!\,(n-r)\,!}=\binom{n}{r}$$

Ex. 2.12.1. *In how many ways can a set of 4 air hostesses be selected from a band of 20 beautiful girls ?*

Solution. Here the order in which the air hostesses are selected is not important. Therefore this is only a case of combination of 20 taking 4 at a time. The answer is

$$\binom{20}{4}=\frac{20.91.18.17}{1.2.3.4}=4845.$$

Ex. 2.12.2. *In a set of 100 birds 50 weigh more than 5 lbs and the rest weigh less than 5 lbs. In how many ways a sample of 10 be selected so that the sample contains 6 birds weighing more than 5 lbs.*

Solution. Here the set of birds weighing more than 5 lbs contains 70 elements. In the sample 6 birds should weigh more than 5 lbs and 4 birds should weigh less than 5 lbs. Selection of 6 birds weighing more than 5 lbs, amounts to the selection of a subset of size 6 from a set of size 70. This can be done in

$\binom{70}{6}$ ways. Similarly 4 birds from 30 birds can be selected in $\binom{30}{4}$ ways. Therefore the total number of ways in which a sample of 10 containing 6 birds weighing more than 5 lbs and 4 birds weighing less than 5 lbs can be taken is

$$\binom{70}{6}\binom{30}{4} \text{ ways.}$$

Ex. 2.12.3. *A deck of 52 playing cards contains 4 different suits of 13 cards each. In how many ways can a hand of 8 cards be selected so that there are 4 clubs, 3 hearts, one diamond and no spades ?*

Solution. By reasoning similar to that of the previous example the answer is $\binom{13}{4}\binom{13}{3}\binom{13}{1}\binom{13}{0}=2658370.$

Ex. 2.12.4. *In how many different ways can 4 people be seated (a) in a row, (b) in a circle ?*

Solution. (*a*) If the people are A, B, C, and D and if they are arranged in a row the first position can be taken by A or B or C or D. The first position can be filled up in 4 ways. Similarly the second position in 3 ways, the third in 2 ways and the 4th in one way. So the total number of arrangements is $4\times3\times2\times1=4\,!$ $=24$.

(*b*) When arranged in a circle all the permutations ABCD, BCDA, CDAB, DABC give the same arrangement. Therefore the number of different seatings $=\frac{4}{4}\,!=3\,!$

In general, when there are k people the answer for (*a*) is $k\,!$ and that for (*b*) is $(k-1)\,!$

Ex. 2.12.5. *How many distinct samples of size r can be selected from a population of size n ?*

Solution. This is only a case of the number of combinations of n things taken r at a time. Hence the answer is $\binom{n}{r}$.

2.13. Stirling's Formula.

According to the definition

$$n\,!=n(n-1)\ldots2.1=1.2.3\ldots n.$$

when

$$n=5 \qquad n\,!=120$$

$$n=8 \qquad n\,!=40320$$

$$n=10 \qquad n\,!=3628800$$

When n becomes large, $n\,!$ becomes very large. For convenience of computation an approximation for $n\,!$ is given by Stirling's formula.

$$n\,! \approx (2\pi)^{1/2}\, n^{n+1/2}\, e^{-n}. \qquad \ldots(2.13.1)$$

where ($\approx$) means approximately equal to, π and e have the usual meaning (*i.e.*, π is approximately=3.1416 and e is approximately equal to 2.71828 where e is the base of natural logarithms).

A closer approximation for n ! is

$$n! \approx (2\pi)^{1/2}\, n^{n+1/2}\, e^{-n+1/12n} \qquad \ldots(2.13.2)$$

Ex. 2.13.1. *Using Stirling's approximation prove that*

$$\binom{2n}{n} \approx \frac{2^{2n}}{(n\pi)^{1/2}}$$

Proof.

$$n! \approx (2\pi)^{1/2}\, n^{n+1/2} e^{-n}$$

$$\binom{2n}{n} = \frac{(2n)!}{n!\,(2n-n)!} = \frac{(2n)!}{(n!)^2}$$

$$(2n)! \approx (2\pi)^{1/2}\,(2n)^{2n+1/2}\, e^{-2n}$$

$$(n!)^2 \approx [(2\pi)^{1/2} n^{n+1/2} e^{-n}]^2 = (2\pi) n^{2n+1} e^{-2n}$$

$$\therefore \quad \frac{(2n)!}{(n!)^2} \approx \frac{(2\pi)^{1/2}(2n)^{2n+1/2} e^{-2n}}{(2\pi) n^{2n+1} e^{-2n}}$$

$$= \frac{2^{2n+1/2}}{(2\pi)^{1/2} n^{1/2}}$$

$$= \frac{2^{2n}}{(\pi n)^{1/2}}.$$

Ex. 2.13.2. *If* $\Gamma(x)$ *(read as gamma* x*) is defined as* $(x-1)$! *for a positive integral value of* x*, show that*

$$\Gamma(x) \approx (2\pi)^{1/2} e^{-x} x^{x-1/2} \text{ when } x \text{ is a positive integer.}$$

Proof.

$$n! \approx (2\pi)^{1/2} n^{n+1/2} e^{-n}$$

$$\Gamma(x) = (x-1)! \approx (2\pi)^{1/2} (x-1)^{(x-1)+1/2} e^{-(x-1)}$$

$$= (2\pi)^{1/2} (x-1)^{x-1/2} e^{-x}.e.$$

$$= (2\pi)^{1/2} x^{x-1/2} \left(1-\frac{1}{x}\right)^{x} e^{-x} \left(1-\frac{1}{x}\right)^{-1/2} e.$$

When x is sufficiently large $\frac{1}{x} \to 0$ (tends to zero) and thereby $\left(1-\frac{1}{x}\right)^{-1/2} \to 1$; $\left(1-\frac{1}{x}\right)^{x} \to e^{-1}$, since $\left(1+\frac{y}{x}\right)^{x} \to e^{y}$ when $x \to \infty$, for a finite y.

$$\therefore \quad \Gamma(x) = (x-1)! \approx (2\pi)^{1/2}.x^{x-1/2} e^{-x}.$$

Comments. In this example x is assumed to be a positive integer. But $\Gamma(x)$ is defined for non-integral values of x also. A reader who is not familiar with limits and gamma functions is advised to read any elementary book on Calculus for 'limits' and J.M.H. Olmsted, Calculus with Analytical Geometry vol. 1 & 2

Appleton-Century-Croft Publishers, New York for Gamma functions.

2·14. The Binomial Coefficients.

Consider the following equations :

$(P_1+P_2)^0=1$

$$(P_2+P_2)^1=P_1+P_2=\binom{1}{0}P_2^0\ P_1^1+\binom{1}{1}P_2^1\ P_1^0$$

$$(P_1+P_2)^2=P_1^2+2P_1P_2+P_2^2=\binom{2}{0}P_2^0\ P_1^2$$
$$+\binom{2}{1}P_2^1\ P_1^1+\binom{2}{2}P_2^2\ P_1^0$$

$$(P_1+P_2)^3=P_1^3+3P_1^2\ P_2+3P_2^2\ P_1+P_2^3=\binom{3}{0}P_2^0\ P_1^3$$
$$+\binom{3}{1}P_2^1\ P_1^2+\binom{3}{2}P_2^2\ P_1^1+\binom{3}{3}P_2^0\ P_1^3$$

In general it may be easily shown that, for any positive integer n,

$$(P_1+P_2)^n=\binom{n}{0}P_2^0\ P_1^n+\binom{n}{1}P_2^1\ P_1^{n-2}+\ldots\ldots$$
$$+\binom{n}{r}P_2^r\ P_1^{n-r}+\ldots\ldots+\binom{n}{n}P_2^n\ P_1^0\ .$$

These types of expansions are called Binomial (of two) expansions and the coefficients of $P_2^0\ P_1^n,\ P_2^1\ P_1^{n-1},\ldots,P_2^n\ P_1^0$ are seen to be $\binom{n}{0},\binom{n}{1},\binom{n}{2},\ldots,\binom{n}{n}$ respectively. These coefficients are called Binomial coefficients. These binomial coefficients may also be constructed from a special arrangement of numbers called the Pascal triangle. The pattern of numbers called Pascal's triangle is given below.

```
                1
            1   .   1
        1   .   2   .   1
    1   .   3   .   3   .   1
1   .   4   .   6   .   4   .   1
.   .   .   .   .   .   .   .   .
.   .   .   .   .   .   .   .   .
  etc.
```

In this arrangement each row starts with one and ends with one. The first row has one element, the second row has two elements etc. Any number in any row is obtained by adding up two consecutive numbers in the preceding row and writing the sum in the next row mid-way between the numbers added up. In this arrangement the second row elements are the coefficients in the expansion of $(P_1+P_2)^1$, the third row elements are the coefficients in the expansion of $(P_1+P_2)^2$ etc. It can be proved by the method of induction that

$$(1+x)^n=1+\binom{n}{1}x+\binom{n}{2}x^2+\ldots\ldots+\binom{n}{r}x^r+\ldots$$

for n a positive integer, negative integer, positive fraction or a negative fraction, where $\binom{n}{r}=\frac{n(n-1)\ldots\ldots(n-r+1)}{1.2\ldots\ldots r}$ and $|x|<1$

When n is a finite positive integer the expansion is valid for any finite x.

Ex. 2.14.1. *Find the total number of samples of sizes 0, 1, 2, 3,......n that can be taken from a population of size n.*

Solution. A sample of size r from a population of size n can be selected in $\binom{n}{r}$ ways. We are asked to find out

$$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\ldots+\binom{n}{n}=\sum_{r=0}^{n}\binom{n}{r}.$$

Consider the expansion

$$(P_1+P_2)^n=\binom{n}{0}P_2^0P_1^n+\binom{n}{1}P_2^1P_1^{n-1}+\ldots+\binom{n}{n}P_2^nP_1^0$$

Put $P_1=P_2=1$ then

$$(1+1)^n=\binom{n}{0}+\binom{n}{1}+\ldots+\binom{n}{r}+\ldots+\binom{n}{n}$$

Hence the answer is 2^n.

Ex. 2.14.2. *Show that* $\binom{m+n}{r}=\sum_{k=0}^{r}\binom{m}{k}\binom{n}{r-k}$

Solution. $(1+x)^{m+n}=(1+x)^m+(1+x)^n$

But
$$(1+x)^{m+n}=\binom{m+n}{0}+\binom{m+n}{1}x+\binom{m+n}{2}x^2+\ldots$$
$$+\binom{m+n}{r}x^r+\ldots$$

$$(1+x)^m=\binom{m}{0}+\binom{m}{1}x+\binom{m}{2}x^2+\ldots\ldots$$

and $$(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \ldots\ldots$$

Comparing the coefficients of x^r on both sides

$$\binom{m+n}{r} = \binom{m}{0}\binom{n}{r} + \binom{m}{1}\binom{n}{r-1} + \ldots + \binom{m}{r}\binom{n}{0}$$

$$= \sum_{k=0}^{r} \binom{m}{k}\binom{n}{r-k}.$$

2·15. The Multinomial Coefficients. The binomial coefficient $\binom{n}{r} = \dfrac{n\,!}{r\,!\,(n-r)\,!}$ (when n is a positive integer) may be written as $\binom{n}{r} = \dfrac{n\,!}{r_1\,!\;r_2\,!}$ where $r_1+r_2=n$, or $\dfrac{n\,!}{r_1!\;r_2\,!}$ may be considered to be the coefficient of $P_1^{r_1}\,P_2^{r_2}$ in the expansion of $(P_1+P_2)^n$ where n is a positive integer, and the whole expansion may be written as

$$\sum_{r_1=0}^{n} \sum_{\substack{r_2=0 \\ r_1+r_2=n}}^{n} \frac{n\,!}{r_1\,!\;r_2\,!}\,P_1^{r_1}\,P_2^{r_2}$$

where the particular Σ notation means that the summation is subject to the condition $r_1+r_2=n$.

This result can be easily generalized to a multinomial expansion for an expression $(P_1+P_2+\ldots\ldots+P_k)^n$. In this expansion the coefficient of $P_1^{r_1}\,P_2^{r_2}\ldots\ldots\ldots P_k^{r_k}$, where $r_1+r_2+\ldots\ldots+r_k=n$, is called a multinomial coefficient. The expansion may be given as

$$\sum_{r_1=0}^{n} \sum_{r_2=0}^{n} \ldots\ldots\ldots \sum_{r_k=0}^{n} \frac{n!}{r_1\,!\;r_2\,!\ldots\ldots r_k\,!}\,P_1^{r_1}\,P_2^{r_2}\ldots\ldots P_k^{r_k}.$$
$$r_1+r_2+\ldots\ldots+r_k=n.$$

Analogous to the binomial coefficient $\dfrac{n\,!}{r_1\,!\;r_2\,!}$, the multinomial coefficient is $\dfrac{n\,!}{r_1\,!\;r_2\,!\;\ldots\ldots r_k\,!}$ where $r_1+r_2+\ldots\ldots+r_k=n$ and n is a positive integer. This may also be considered to be the number of ways of getting r_1, P_1's, r_2, P_2's......r_k, P_k's, such that $r_1+r_2+\ldots\ldots+r_k=n$. Another notation for the multinomial coefficient $\dfrac{n\,!}{r_1\,!\,r_2\,!\;\ldots\ldots r_k}$ is $\dbinom{n}{r_1,\,r_2,\,\ldots\ldots,\,r_k}$.

It is easily seen that

$$\frac{n\,!}{r_1\,!\;r_2\,!\;\ldots\ldots r_k\,!} = \binom{n}{r_1}\binom{n-r_1}{r_2}\binom{n-r_1-r_2}{r_3}\ldots\ldots\binom{r_k}{r_k}$$

that is, the multinomial coefficient may also be considered to be the number of ways in which a set of n elements can be arranged into an ordered set of k subsets having $r_1, r_2, \ldots\ldots r_k$ elements respectively, where $r_1+r_2+\ldots\ldots+r_k=n$. It is also seen to be the number of permutations of n objects taking n at a time, in which r_1 are of one type, r_2 of a second type, $\ldots\ldots r_k$ are of a k^{th} type, where $r_1+r_2+\ldots\ldots+r_k=n$.

Ex. 2.15.1. *A coin is tossed 10 times. What is the total possible number of ways in which 4 heads and 6 tails can come up ?*

Solution. This can be obtained from a binomial expansion of the form $(P_1+P_2)^{10}$. The various coefficients in the expansion give the various possibilities of zero P_2 and 10 P_1, one P_2 and 9 P_1, ……, ten P_2 and zero P_1. Therefore the coefficient of $P_1^6\ P_2^4$ gives the required number of ways in which 4 heads and 6 tails can come up and is

$$=\binom{10}{6}=\binom{10}{4}$$

$$=\frac{10\,!}{4\,!\,6\,!}=\frac{10\,.\,9\,.\,8\,.\,7}{1\,.\,2\,.\,3\,.\,4}$$

$$=210.$$

Ex. 2.15.2. *A die is rolled 12 times. In how many ways can one get two 1's, two 2's, one 3, four 4's, two 5's and one 6 ?*

Solution. This is similar to the previous example and the required number is the coefficient of $P_1^2\ P_2^2\ p_3^1\ p_4^4\ p_5^2\ p_6^1$ in the expansion of $(p_1+p_2+p_3+p_4+p_5+p_6)^{12}$. The answer is,

$$\frac{12\,!}{2\,!\,2\,!\,1\,!\,4\,!\,2\,!\,1\,!}=\binom{12}{2,\ 2,\ 1,\ 4,\ 2,\ 1}$$

$$=2494800.$$

Exercises

2·1. Find the numbers of vectors and subsets that can be obtained from the sets,

(a) $A=\{0, 1, 2, 3\}$,

(b) $B=\{-1, 5, 0\}$.

2·2 Ku. Nalini can make pies with coconut, pumpkin and apple, with soft or hard crusts, and with sugar contents 1%, 2% or 3%. How many distinct types of pies can she make ?

2·3. A four letter word is to be made with the letters of the Latin alphabet. How many distinct words can be made if the first letter of the word is not A ?

2·4. In how many different ways can 4 people be seated

(i) in a row,

(2) in a circle,

(3) in a row of 6 seats ?

2·5. How many distinct words can be made using all the letters of the word panamanian ?

2·6. A student's council has 5 members. How many different ways can both a programme committee and a reception committee, each with two members, be made, if

(*a*) nobody is allowed to sit on both the committees,

(*b*) with no restriction ?

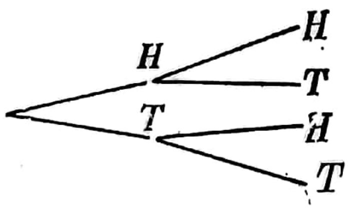
2·7. In how many ways can

(1) an athltic team of 5 be selected from a set of 30 sportsmen,

(2) a team of 5 be selected with Shri Kumar of the set in the selected team ?

2·8. A secretary puts letters addressed to 4 people into the envelopes which are already addressed to the 4 people. In how many ways can she put the letters in the envelopes, so that nobody receives the letter addressed to him ? (This problem is often called the matching problem.). Prove that in the general case when there are k envelopes and k letters the answer is $\frac{1}{2\,!}-\frac{1}{3\,!}+\,\cdots\cdots+\frac{(-1)^k}{k\,!}$.

2·9. If the outcomes head and tail, when a coin is tossed once, are denoted by H and T, we may construct a diagram which may be called a tree diagram to compute the number of possible outcomes when a coin is tossed a number of times. This diagrammatic representation is helpful in many problems. Such a diagram for the experiment of throwing a coin twice is given below.



Using a tree diagram determine the total number of outcomes when a die is rolled three times.

2·10. In how many ways can a set of 13 cards be selected from 52 cards such that it contains 4 hearts, 5 clubs, 4 diamonds and a spade ?

2·11. Ku. Ragini knows 5 different Bharat natya (an Indian classical dance). In how many different ways can she give the performance for two occasions if

(1) she does two dances for each performance and different dances for different performances,

(2) two dances for each performance but she starts with the same dance for every performance ?

2·12. In a class there are 100 students out of which 20 are highly intelligent. In how many ways a committee of 20 be formed so that the committee contains 4 highly intelligent ones ?

2·13. In how many ways can 6 halwa bars (an Indian sweet preparation) be distributed among 4 children so that

(1) any child can receive any number of halwa bars,

(2) there should be at least one halwa bar for every child ?

Show that the general formula for the number of ways in which r indistinguishable articles can be assigned into n cells is $\binom{n+r-1}{r}$. (This is known as the occupancy problem).

2·14. Show that

(a) $\sum_{r=0}^{n} (-1)^r \binom{n}{r}=0$;

(b) $\sum_{r=0}^{n} \binom{n}{r}^2=\binom{2n}{n}$;

(c) $\sum_{r=0}^{2k} (-1)^r \binom{n}{r}\binom{n}{2k-r}=(-1)^k \binom{n}{k}$ for $0\leqslant 2k\leqslant n$;

(d) Evaluate $\binom{-1/2}{2}$ and $\binom{3/4}{4}$;

(e) $\binom{2n}{n}=(-1)^n\, 2^{2n} \binom{-1/2}{n}$.

2·15. By using Stirling's formula if necessary,

(a) Show that $n\,!\approx(2\pi)^{1/2}\ (n+1/2)^{n+1/2}\ e^{-(n+1/2)}$,

(b) approximate 15 ! and find the percentage error in the approximation.

2·16. In how many ways can 6 rolls of a die yield 2 ones, 3 twos, one 3 and no 4's, 5's and 6's ?

2.2. ALGEBRA OF SETS

Mathematical operations analogous to the simple mathematical operations of addition, multiplication etc. on numbers, can be defined on sets.

2.21. Union of Sets. Union or logical sum of two sets A and B, which is written as A$\cup$B (A union B), is defined as the set of elements which belong to at least one of the sets A and B or the set of elements which belong to A or B or both. Other notations are A+B, A$\dot{+}$B etc.

Ex. 2.21.1. *Let* $A=\{1, 2, 3, 4\}$; $B=\{1, -1, 0, 4\}$

Here 1, 2, 3, 4 belong to A, 1, —1, 0, 4 belong to B and further 1 and 4 belong to both A and B. Therefore

$$A\cup B=\{1, 2, 3, 4, -1, 0\}$$

Comments. The numbers 1, 2, 3, 4, -1, 0 belong to atleast one of the sets A and B. Here the order in which the elements are written and the magnitudes of the elements have no importance. For example if A$=\{5\}$ and B$=\{4\}$ then A$\cup$B$=\{5, 4\}$ $=\{4,5\}\neq\{5+4\}=\{9\}$. The operator 'union of two sets' is a binary operation, that is an operation connecting two mathematical objects.

Ex. 2.21.2. *$A=\{5$, Ku. Sarojam, Mr. Iceberg, $\theta\}$; $B=\{4, 3, -1$, Tajmahal, Mr. Fox, Shri. Ajaya$\}$ then $A \cup B=\{5$, Ku. Sarojam, Mr. Fox, Mr. Iceberg, θ, 4, 3, -1, Tajmahal. Shri. Ajaya$\}$.*

Comments. The elements of a set need not be always numbers in order to define the operation 'union' on them.

Ex. 2.21.3. *Let A and B be the spaces enclosed by the closed curves α and β in Fig. 2.1 respectively then $A \cup B$ is the shaded region.*

Comments. Representation of sets by diagrams as shown in Fig. 2.1 is called representation by Venn diagrams. This will enable the reader to grasp quickly the ideas of sets and union of sets etc. In a Venn diagrammatic representation a set is represented by the space enclosed by a closed curve thereby the elements of the set are symbolically represented as points in this space. This does not mean that the elements of a set can always be represented as geometrical points in a two dimensional space. The representation is purely symbolic. For example A and B can be the sets A and B in Ex. 2.21.2. In this case the space α (alpha) has 4 points and β (beta) has 6 points or 6 elements.

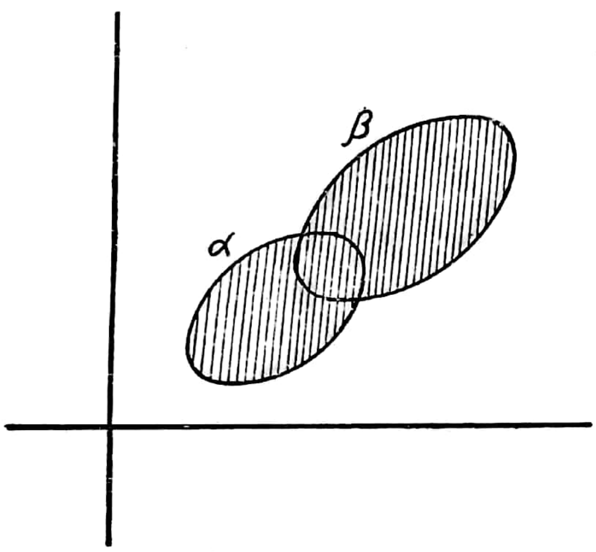


Fig. 2.1.

Ex. 2.21.4. *$A=$the set of students who take a particular mathematics course in a particular university at a particular time. $B=$ the set of students who take another mathematics course in that university at that time. Then $A \cup B$ is the set of students who take atleast one of the courses under consideration.*

Comments. If all the students in one course are also students in the second course and if no other students take the second course then A and B are the same, and hence evidently $A \cup B=A=B$. We can easily prove the general result that $A \cup A=A$ where A is any set.

Ex. 2.21.5. *$A=$the set of rose flowers in Brindawan garden at a particular time. $B=$the set of pink rose flowers in Brindawan at that time. Then $A \cup B$ is the set of rose flowers in Brindawan which are pink or not.*

Comments. Here evidently $B \subset A$ and hence $A \cup B=A$. If there are no pink roses then $B=\phi$ and therefore $A \cup \phi=A$. These results hold good for any set A and for any subset B of A.

The above definition of union of two sets may be extended

to a collection of n sets. If $A_1 \ldots, A_n$ is a collection of n sets then the union of $A_1, \ldots, A_n$ (that is, $A_1 \cup A_2 \cup \ldots \cup A_n = \bigcup_{i=1}^{n} A_i$) is that set of elements which belong to atleast one of the sets $A_1, \ldots, A_n$.

Ex. 2.21.6. *Let* $A_1 = \{1, 2, 3\}$; $A_2 = \{0, -1, 2\}$; $A_3 = \{7, 0, 1/2\}$; $A_4 = \{5, 1\}$ *then* $A_1 \cup A_2 \cup A_3 \cup A_4 = \{1, 2, 3, 0, -1, 7, 1/2, 5\}$.

Comments. From the above examples it is evident that for any sets A, B, C, (1) $A \cup B = B \cup A$ (that is, the binary operation 'union' is commutative. An operation P connecting two mathematical objects a and b is said to be commutative if $aPb = bPa$. For example the simple 'addition' is commutative whereas 'subtraction' is not commutative since $a+b=b+a$ but $a-b \neq b-a$ in general), (2) $A \cup B \cup C = B \cup C \cup A = C \cup B \cup A$, (3) $A \cup (B \cup C) = (A \cup B) \cup C$ (that is, the operation 'union' is associative or in other words it does not make any difference whether $B \cup C$ is done first or $A \cup B$ is done first. The simple operation 'subtraction' is not associative since $a-(b-c) \neq (a-b)-c$ where a, b, c are real numbers).

Ex. 2.21.7. *Consider an experiment of throwing a coin three times. Let the occurrence of a head be denoted by 1 and that of a tail by zero. Then the outcomes set may be given as,*

$S = \{(0, 0, 0), (0, 0, 1,), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$. *Let A be the event, of obtaining a total of 2, B be the event of obtaining a total 1, and C be the event of getting a total 0, then,* $A = \{(0, 1, 1), (1, 0, 1), (1, 1, 0)\}$; $B = \{(0, 0, 1), (0, 1, 0), (1, 0, 0)\}$, $C = \{(0, 0, 0)\}$,

Then $A \cup C = \{(0, 1, 1), (1, 0, 1), (1, 1, 0), (0, 0, 0)\}$ *gives the event of getting a total 0 or 2.*

$$A \cup B \cup C = \{(0, 1, 1,), (1, 0, 1), (1, 1, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (0, 0, 0)\}$$

gives the event of getting a total of 0 or 1 or 2.

Comments. In this example events A and B or the events of getting totals of 1 and 2 can not happen simultaneously. A statement of simultaneous occurrence of A and B, in this experiment, has no meaning. So also the events B and C, and also A and B and C can not happen simultaneously.

Ex. 2.21.8. *Consider an experiment of rolling one die once. The outcome set may be given as the points shown in Fig. 2.2.*

Let A be the event of rolling 1 or 3 and B be the event of rolling 4, then $A \cup B$ is the event of rolling 1 or 3 or 4.

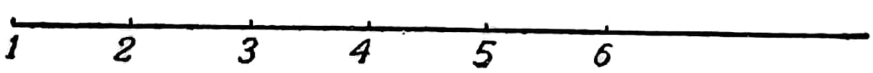


Fig. 2·2.

Comments. If two dice are rolled the outcome set may be represented as a set of points in a two dimensional space etc.

2.22. Intersection of two Sets. The set of elements which belong to both A and B is called the intersection of the sets A and B and is usually denoted by $A \cap B$ (A intersection B), AB etc. Intersection is also called logical product. Intersection is analogous to the operation 'multiplication' but it is different from multiplication.

Ex. 2.22.1. *Let* $A=\{2, 4, -7\}$

$B=\{2, 5\}$

then $A \cap B=\{2\}$

Comments. If $A=\{6\}$ and $B=\{7\}$ then $A \cap B \neq \{6 \times 7\}=\{42\}$

Ex. 2.22.2. *A={Ajayan, Ku. Lalitha, 8}*

B={Miss Cute, Ku. Lalitha}

then *A∩B={Ku. Lalitha}*

Ex. 2.21.3. *A—the set of all students in a class*

B—the set of all female students in the class.

Then $A \cap B=B=$ *the set of all female students in the class.*

Comments. Here $B \subset A$ and therefore $A \cap B=B$. If there are no female students then $B=\phi$ and evidently $A \cap \phi=\phi$. If all the students are female students then $B=A$ and therefore $A \cap A=A$. So we can easily prove the following results that for any set A, $A \cup A=A$, $A \cap A=A$, $A \cup \phi=A$, $A \cap \phi=\phi$ and if $B \subset A$ then $A \cup B=A$ and $A \cap B=B$.

Ex. 2.21.4. *Let A and B be the sets as shown in the Venn diagram ; then* $A \cap B$ *is the shaded region.*

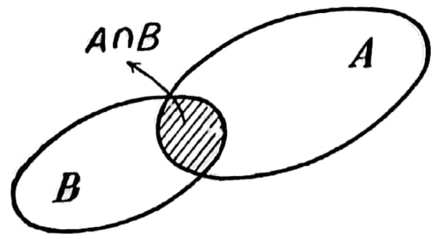


Fig. 2·3.

Comments. From the above examples it is evident that $A \cap B=B \cap A$. The same definition of intersection may be extended to a number of sets. The intersection of the sets $A_1, A_2, \ldots, A_n$ (denoted by $A_1 \cap A_2 \cap \ldots \cap A_n=\bigcap_{i=1}^{n} A_i$) may be defined as the set of elements which belong to all the sets $A_1, A_2, \ldots A_n$.

2.23. Disjoint Sets and Mutually Exclusive Events. A number of sets $A_1, A_2, \ldots, A_n$ are said to be pairwise disjoint if $A_i \cap A_j = \phi$ for $i \neq j$ for all i and j. *i.e.*, $A_1 \cap A_2 = \phi$, $A_1 \cap A_3 = \phi$, $A_2 \cap A_3 = \phi$ etc. If $A_1, A_2, \ldots A_n$ denote disjoint subsets of an outcome set then the events $A_1, A_2, \ldots A_n$ are said to be mutually exclusive. *i.e.*, two events A and B are mutually exclusive if $A \cap B = \phi$. Here A and B have no element in common or the occurrence of one excludes the occurrence of the other.

Ex. 2.23.1. *Suppose that the points marked in Fig. 2·4 represent an outcome set ; then the two shaded regions represent two mutually exclusive events.*

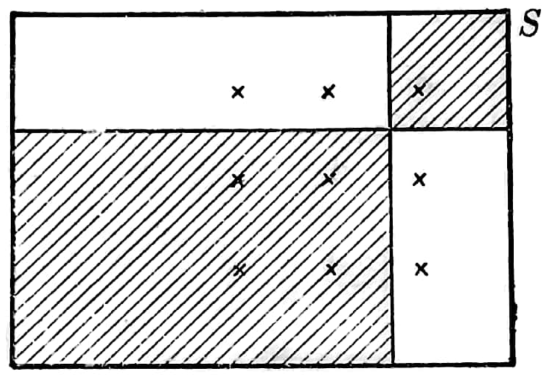


Fig. 2·4.

Comments. Here the shaded portions are not important, but only the points in the shaded parts are important. The two parts have no common points and hence the corresponding events are mutually exclusive.

Ex. 2.23.2. *Consider an experiment of throwing a coin three times. The outcome set is given by*

$$S = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}.$$

The event A of getting a total 2 is given by

$$A = \{(0, 1, 1), (1, 0, 1), (1, 1, 0)\}.$$

The event B of getting 1 is given by

$$B = \{(0, 0, 1), (1, 0, 0), (0, 1, 0)\}$$

$A \cap B = \phi$ since A and B have no element in common. A and B are mutually exclusive or when we get a total 1 we cannot get a total 2 at the same time. The two events cannot occur simultaneously.

Ex. 2·23·3. *Let the outcome set, S and the events A, B, C, D be as shown in Fig. 2·5 then A, B, C, D are mutually exclusive.*

Comments. Here S is partitioned into disjoint sets or this is a disjoint partition of the outcome set. Evidently.

$$A \cup B \cup C \cup D = S, \ A \cap B = \phi, \ A \cap C = \phi, \ A \cap D = \phi, \ B \cap C = \phi \text{ etc.}$$

2.24. Complement of a set A or non-occurrence of an event A. The set of elements which do not belong to A may be

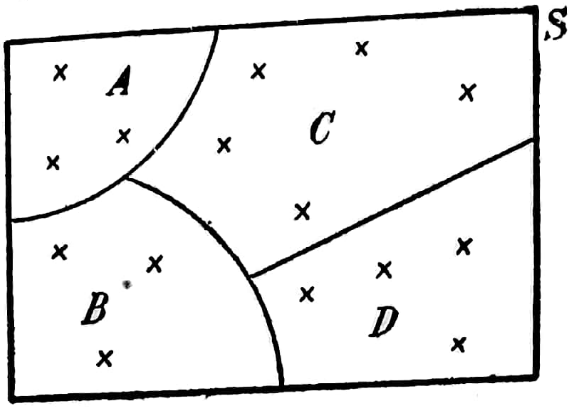


Fig. 2·5

called the complement of A and is usually denoted by $\bar{A}$. Other notations are A', A, A^{e} etc. Therefore the complement of A with respect to B is the set of points in B which do not belong to A and is usually denoted by B—A.

Ex. 2.24.1. *Let the sets A and S be as shown in Fig. 2·6 then A is the shaded portion in Fig. 2·6.*

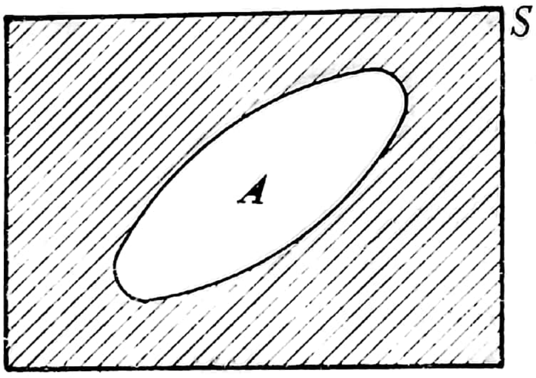


Fig. 2·6.

Comments. The set of all points in the shaded portion gives the non-occurrence of the event A, if S is an outcome set. In this case $A \subset S$ and also $\bar{A} \subset S$.

Ex. 2.24.2. *Let A, B, S be given as in Fig. 2·7, then the various complements with respect to the outcome set S and with respect to events A and B are marked in Fig. 2·7.*

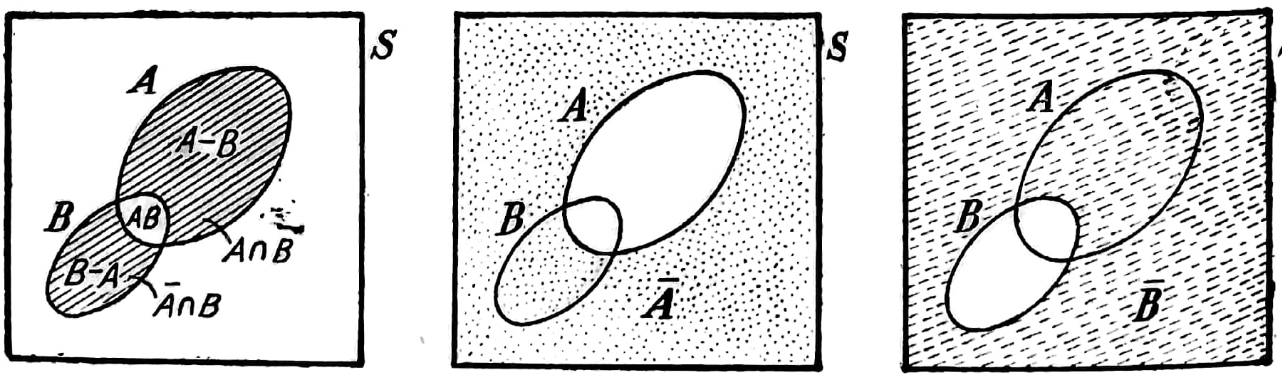


Fig. 2·7.

Ex. 2.24.3. *Let* $S=\{x \mid 1 \leqslant x \leqslant 6\}$ *be the outcome set*

$A=\{x \mid 1 \leqslant x \leqslant 3\}$ *be an event.*

$B=\{x \mid 2 \leqslant x \leqslant 4\}$ *be an event.*

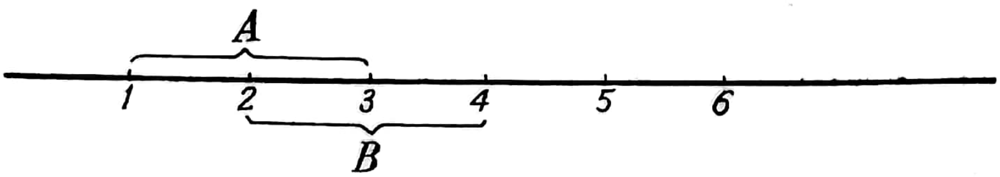


Fig. 2·8.

Then $\bar{A}=\{x \mid 3<x \leqslant 6\}$ (non-occurrence of the event A)

$\bar{B}=\{x \mid 1 \leqslant x<2, \text{ or } 4<x \leqslant 6\}$ (non-occurrence of the event B)

$A \cup B=\{x \mid 1 \leqslant x \leqslant 4\}$ (event of the occurrence of at-least one of the events A and B).

$A \cap B=\{x \mid 2 \leqslant x \leqslant 3\}$ (event of simultaneous occurrence of A and B)

$A \cup \bar{A}=\{x \mid 1 \leqslant x \leqslant 6\}=S$ (event that either A occurs or A does not occur. This is evidently a sure event).

Comments. It may be easily verified that $A \cup S=S$, $A \cap S=A$, $A \cup (A \cap B)=A$. The following results are generally true. If A is any event (1) $A \cup \bar{A}=S$, $A \cup S=S$, $A \cap S=A$.

Ex. 2.24.4. *Consider an experiment of throwing a coin three times. The outcome set S may be given as,*

$S=\{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1\ 0, 1), (1, 1, 0), 1, 1, 1)\}$

Let A be an event of getting a total 2 and B be the event of getting a total 1, then

$A=\{(0, 1, 1), (1, 0, 1), (1, 1, 0)\}$

$B=\{(0, 0, 1), (0, 1, 0), (1, 0, 0)\}$

$\bar{A}$=Set of all vectors other than the vectors in A=the event of getting a total 0 or 1 or 3, or $\bar{A}$ is the event of getting a total not equal to 2.

Comments. It can be easily verified that $A \cap B=\phi$.

Ex. 2.24.5. *Let A, B, C be the events as shown in the Venn diagram in Fig. 2·9, then* $A \cap B \cap C$ *is given by the shaded region.*

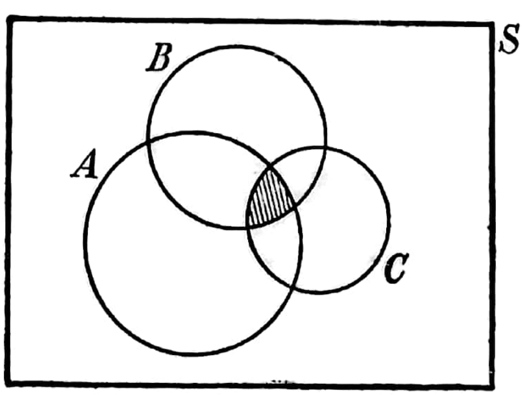


Fig. 2·9.

Comments. The following results may be easily verified from Fig. 2·9.

(1) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,

(2) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,

(3) $\overline{A \cup B} = \bar{A} \cap \bar{B}$,

(4) $\overline{A \cap B} = \bar{A} \cup \bar{B}$,

(5) $\overline{A \cup B \cup C} = \bar{A} \cap \bar{B} \cap \bar{C}$,

(6) $\overline{A \cap B \cap C} = \bar{A} \cup \bar{B} \cup \bar{C}$.

and all these results can be easily generalized to any finite number of sets.

From the above examples a correspondence between sets, subsets, etc., to outcome set, events etc., is seen. These may be summarized and given as follows.

	Events	**Outcome sets**
1.	Elementary event	Point belonging to an outcome set
2.	Event	Subset of an outcome set
3.	Sure event	Whole of the outcome set
4.	Impossible event	Null set which is evidently a subset of the outcome set
5.	Non-occurrence of an event	Complement of a subset of an outcome set
6.	Mutually exclusive events	Disjoint subsets of an outcome set
7.	Occurrence of atleast one of the events A and B	$A \cup B$, where A and B are subsets of an outcome set
8.	Simultaneous occurrence of the events A and B	$A \cap B$ where A and B are subsets of the outcome set

Exercises

2·17. If A and B are two sets such that $A \cup B = \{2, 5, 0, -1\}$ and $A \cap B = \{2\}$, find A and B. Are they unique ?

2·18. In an experiment of throwing a coin 3 times find (*a*) three events which are mutually exclusive, (*b*) the complement of the event of getting exactly one head.

2·19. If $S = \{0, 1, 2, 3, 4, 5, 6\}$, $A = \{0, 1, 2, 3\}$, $B = \{3, 4, 5\}$, find (1) $\bar{A}$, (2) $A \cup B$, (3) $A \cap B$, (4) complement of B with respect to A.

2·3. Functions. The reader may be familiar with point set functions or functions of the type $y = x^3 + x + 2$, $y = \sin x$, etc. where in general we get a curve in a two dimensional space. The

relationship between the variables x and y, when represented in a graph, is given by a curve as shown in Fig. 2·10.

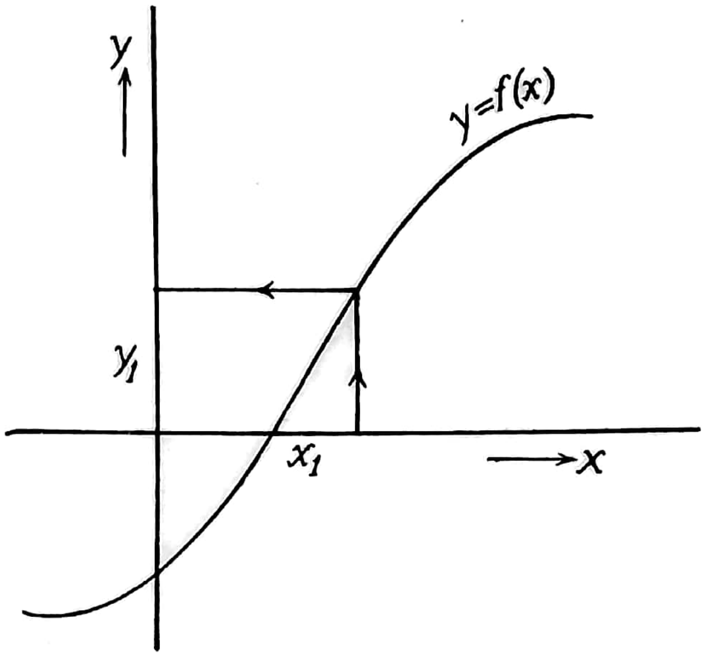


Fig. 2·10.

Corresponding to a point x_1 on the X-axis we get y_1 on the Y-axis. This curve $y=f(x)$ can be considered to be a correspondence between the points on the X-axis and the points on the Y-axis where the correspondence or the rule which defines the correspondence is given by $y=f(x)$. We will generalize this notion and will define a function or a mapping as a correspondence between the elements of two sets.

Definition. A function or a mapping is a correspondence between the elements in the sets X and Y such that corresponding to an element in X there is a unique element in Y. This mapping can be written as

$$X \xrightarrow{f} Y$$

For convenience we may write the correspondence as $y=f(x)$ where x is any element in X and y is the corresponding element in Y. By this definition a number of points in X may be mapped to the same point in Y but corresponding to any point in X there is only one point in **Y**.

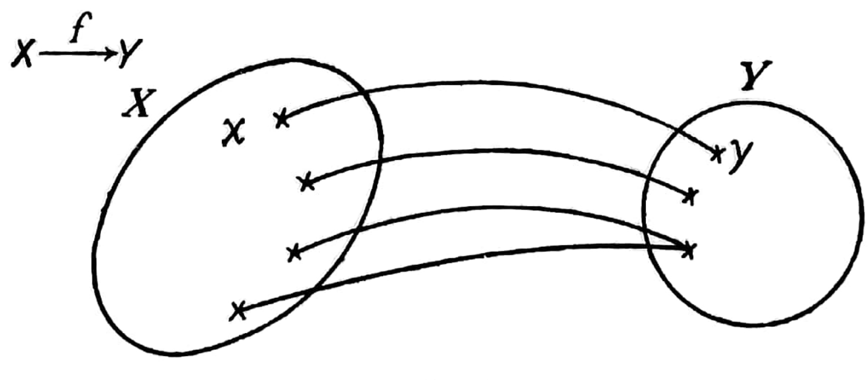


Fig. 2·11.

X is called the domain of the function and Y is called the range of the function.

Ex. 2.3.1. *Consider the set X and Y where*

X	Y
1	2
2	5
-3	10
6	37
3	10

In these two sets, if any element in X is denoted by x and the corresponding element in Y is denoted by y then the correspondence may be written as $y=x^2+1$.

Comments. Sometimes it is possible to express the correspondence in known functions [*i.e.*, the form of $f(x)$ is known], like $y=2x^3+3x+5$, $y=\sin x$, $y=\log x+\dfrac{1}{x}$ etc.

In the function $y=x^2+1$ if x is defined on the set $R=\{x \mid -\infty < x < \infty\}$ then the domain of the function is the real line R and the range is evidently the line $S=\{y \mid 1 \leqslant y < \infty\}$. Here both the domain and the range are sets of numbers.

Ex. 2.3.2. *Consider the sets X and Y where the x's and y's represent the height and weight measurements of the students in a class.*

X	Y
x_1	y_1
x_3	y_2
....	....
x_n	y_n

For example x_1 is the height of one student and y_1 denotes his weight etc.

Comments. In this case we may not be able to find out correspondence in a functional form like $y=x+2$ or $y=x^2-x+3$ etc. Here the elements in the sets X and Y are numbers. According to the definition X and Y need not be sets of numbers. The elements in X and Y can be other objects or sets etc. Even though we denote the mapping by $y=f(x)$ the functional form of $f(x)$ need not be always known.

Ex. 2.3.3.

X	Y
$A_1=\{x \mid 0 \leqslant x \leqslant 1\}$	1
$A_2=\{x \mid 4 \leqslant x \leqslant 7\}$	3
$A_3=\{x \mid 6 \leqslant x \leqslant 10\}$	4
$A_4=\{x \mid -2 \leqslant x \leqslant 0\}$	2

Comments. Evidently the elements in X are sets (intervals) and the elements in Y are numbers (namely the lengths of the intervals).

Ex. 2.3.4. *Consider an experiment of tossing a coin three times. The elements in the outcome set are given below :*

$$V_1=(H, H, H),\ V_2=(H, H, T),\ V_3=(H, T, H),\ V_4=(H, T, T),$$
$$V_5=(T, T, T),\ V_6=(T, T, H),\ V_7=(T, H, T),\ V_8=(T, H, H).$$

where H denotes a head and T denotes a tail. Let X be the set of events of getting one head, 2 heads and 3 tails respectively. Let Y denote the number of outcomes or the number of elementary events in the events mentioned above, then,

X	Y
$A_1=\{V_4, V_6, V_7\}$	3
$A_2=\{V_2, V_3, V_8\}$	3
$A_3=\{V_5\}$	1

Comments. In examples 2.3.3 and 2.3.4 the elements in X are sets and those in Y are numbers. The domain of the function is a set of sets and the range is a set of real numbers. Such functions are called set functions.

2·31. Set Functions. A function whose domain is a set of sets and whose range is a set of real numbers may be called a set function. In the following discussions we are interested only in set functions. In example 2·3·3 it is seen that $f(A_1 \cup A_2)=f(A_1)+f(A_2)$. In other words if x is an interval and if y is the length of the interval x, then evidently the length of the union of two disjoint intervals is the sum of the lengths of the intervals. In the same example

$$f(A_2 \cup A_3)=f(4 \leqslant x \leqslant 10)=6$$

But $$f(A_2)+f(A_3)=3+4=7.$$

Here it is easily seen that $f(A_2 \cup A_3)=f(A_2)+f(A_3)-f(A_2 \cap A_3)$

i.e., $f(4 \leqslant x \leqslant 10)=f(4 \leqslant x \leqslant 7)+f(6 \leqslant x \leqslant 10)-f(6 \leqslant x \leqslant 7)$
$$=3+4-1=6.$$

2·32. Additivity. In a set function if $f(A_1 \cup A_2 \cup \dots \cup A_n)=f(A_1)+f(A_2)+\dots+f(A_n)$ when $A_1, A_2, \dots\dots, A_n$ are disjoint in the sense that the intersection of any two is a null set ($A_i \cap A_j=\phi$ for all i and j, $i \neq j$) then the set function is called additive. In the examples 2·3·3 and 2·3·4 the set functions are additive. If a set S is partitioned into a countable number of disjoint sets $A_1, A_2, \dots$ and if a set function defined on S is such that

$$f(A_1 \cup A_2 \cup \dots \cup A_n \dots)=f(A_1)+f(A_2)+\dots$$

then the function is called totally additive. (If we can number the partitions in S by the natural numbers 1, 2, 3...then the sets

$A_1, A_2, \ldots$ are called a countable partition of S or if in a set there is a correspondence between the elements and the natural numbers, 1, 2, 3,......then the set is said to have a countable number of elements). The total additivity condition may be written as

$$f\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} f(A_i) \text{ where } A_i \cap A_j=\phi,\ \forall\ i \text{ and } j,\ i \neq j$$

Here $\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup A_3 \cup \ldots$ and $\forall$ means 'for all'.

Total additivity is a property concerning the nature of the mapping. For example the mapping intervals→lengths of intervals is of the type that when a line segment PQ is divided into a countable number of disjoint segments the sum of the lengths of these segments equals the length of the line segment PQ.

2.33. Measure. A set function which is non-negative and totally additive is called a measure. This means that if $f(S)$ is a set function defined on S and if $f(S)$ is a measure it should satisfy the following conditions.

$$f(S) \geqslant 0$$

$$f\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} f(A_i) \text{ where } A_i \cap A_j=\phi,\ \forall\ i \text{ and } j,\ i \neq j$$

$$f(A_i) \geqslant 0 \text{ for all } A_i,\ i=1, 2, 3,\ldots$$

In examples 2·3·3 and 2.3.4 the set functions are measures, because they satisfy all the conditions above.

2.34. Probability Measure. If in a measure the total measure equals unity or $f(S)=1$ then the measure $f(S)$ is called a probability measure. This definition implies that if S is an outcome set then $f(S)=1$. A probability measure is usually denoted by P(S). If $A_1, A_2, \ldots, A_n$ are a disjoint partition of the outcome set S then $A_1, A_2, \ldots, A_n$ are mutually exclusive events, and

$$S=A_1 \cup A_2 \cup \ldots \cup A_n = \bigcup_{i=1}^{n} A_i.$$

Then $P(S)=P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i)=1.$

This shows that the probability of an event satisfies the conditions.

(1) $0 \leqslant P(A) \leqslant 1$

(2) $P(S)=1$

(3) $P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i)$ when $A_i \cap A_j=\phi,\ \forall\ i$ and $j,\ i \neq j$.

Therefore the probability of an event A may be defined by the three axioms (1), (2) and (3). These axioms give the properties or the desirable qualities of the probability of an event A, but they do not give any clue to, how to assign a probability to an event A or how to evaluate the probability of an event A.

Let S be an outcome set. Let A, B, C be three mutually exclusive events such that $A \cup B \cup C = S$. According to the axioms or assumptions or postulates the probabilities of the events A, B, C, should satisfy the conditions,

$$0 \leqslant P(A) \leqslant 1, \; 0 \leqslant P(B) \leqslant 1, \; 0 \leqslant P(C) \leqslant 1$$

$$P(S) = P(A \cup B \cup C) = P(A) + P(B) + P(C) = 1.$$

The following are some of the possibilities :

$$P(A) = 1/3 \quad P(B) = 1/3 \quad P(C) = 1/3$$
$$P(A) = 1/6 \quad P(B) = 1/2 \quad P(C) = 1/3$$
$$P(A) = 0{\cdot}2 \quad P(B) = 0{\cdot}7 \quad P(C) = 0{\cdot}1 \text{ etc.}$$

Comments. There are a number of ways in which we can assign probabilities to A, B, and C, according to the axioms. In order to determine uniquely the probability of an event A we need some more considerations. To some extent it is possible to determine the probability of an event A by considering the experimental conditions, symmetry, past experience etc.

Theorem 2.3.1.

$$P(\phi) = 0$$

Proof. If S is the outcome set $S \cup \phi = S$ and further S and ϕ are mutually exclusive.

$$\therefore \quad P(S \cup \phi) = P(S)$$

i.e., $$P(S \cup \phi) = P(S) + P(\phi) = P(S) \Rightarrow P(\phi) = 0.$$

Comments. The probability of an impossible event is zero, but the converse that, if the probability of an event is zero the event is impossible, need not be true. This point will be discussed later.

Theorem 2.3.2.

$$P(\bar{A}) = 1 - P(A)$$

Proof. If S is the outcome set and if A is an event

$A \cup \bar{A} = S$ and further A and $\bar{A}$ are mutually exclusive.

$\therefore \quad P(S) = P(A \cup \bar{A}) = P(A) + P(\bar{A}).$

But $P(S) = 1 \Rightarrow 1 = P(A) + P(\bar{A})$ or $P(\bar{A}) = 1 - P(A)$.

For example if the probability of the occurrence of an event A is $0{\cdot}4$ then the probability of non-occurrence of the event A is $1 - 0{\cdot}4 = 0{\cdot}6$.

Theorem. 2·3·3.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Proof.

$$A \cup B = A \cup (\bar{A} \cap B)$$

and A and $\bar{A} \cap B$ are mutually exclusive.

$$\therefore \quad P(A \cup B) = P(A) + P(\bar{A} \cap B)$$
$$= P(A) + [P(\bar{A} \cap B) + P(A \cap B)] - P(A \cap B)$$

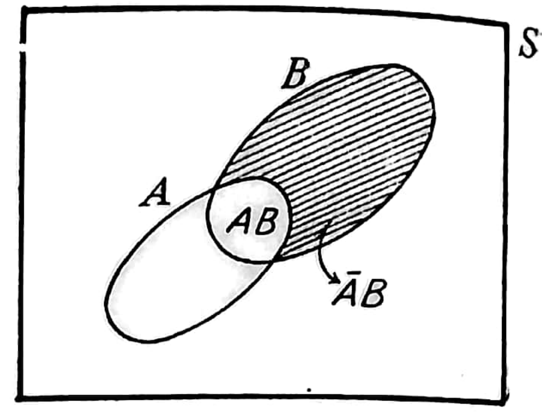


Fig. 2·12.

But $\bar{A} \cap B$ and $A \cap B$ are mutually exclusive.

$$\therefore \quad P(\bar{A} \cap B) + P(A \cap B)$$
$$= P[(\bar{A} \cap B) \cup (A \cap B)]$$
$$= P[(\bar{A} \cup A) \cap B]$$
$$= P[S \cap B]$$
$$= P(B)$$

$$\therefore \quad P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Ex. 2.34.2. *Suppose that the probability that a woman entering a shop buys chewing gum is 0 80. the probability that she buys sleeping pills is 0·70, the probability that she buys chewing gum and sleeping pills is 0·55. What is the probability that a woman entering the shop buys chewing gum or sleeping pills or both ?*

Solution. Let A be the event that a woman entering a shop buys chewing gum, and let B be that of buying sleeping pills, then $P(A) = 0{\cdot}80$, $P(B) = 0{\cdot}70$ and $P(A \cap B) = 0{\cdot}55$.

Therefore, $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0{\cdot}80 + 0{\cdot}70 - 0{\cdot}55$
$= 0{\cdot}95$.

Comments. It may be noticed that if $A \cap B = \phi$ then $P(A \cap B) = 0$ and $P(A \cup B) = P(A) + P(B)$.

Ex. 2.34.3. *Suppose the probability that, a man selling curios at Kanya Kumari in the southern tip of India will have a customer on a Sunday afternoon who will buy a decorated sea shell is 0·80, that he will buy a sea shell chain is 0·40, that he will buy a decorated shell and a sea shell chain is 0·30, what is the probability that the shopkeeper has a customer who will buy at least one of the items, a decorated shell and a sea shell chain ?*

Solution. Let A and B denote the events that the customer will buy a decorated sea shell and a sea shell chain respectively, then $P(A) = 0{\cdot}80$, $P(B) = 0{\cdot}40$ and $P(A \cap B) = 0{\cdot}30$. We are asked to find out $P(A \cup B)$.

But $\quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= 0{\cdot}80 + 0{\cdot}40 - 0{\cdot}30 = 0{\cdot}90$.

Exercises

2·20. Give three examples each for

(1) an additive set function,

(2) a measure,

(3) a probability measure.

2·21. If A, B, C are a disjoint partition of the outcome set S then can the following measures be probability measures ?

(*a*) $P(A)=0{\cdot}5,\ P(B)=0{\cdot}3,\ P(C)=0{\cdot}2$

(*b*) $P(A)=0{\cdot}5,\ P(B)=0{\cdot}8,\ P(C)=0{\cdot}7$

(*c*) $P(A)=0{\cdot}7,\ P(B)=0{\cdot}3,\ P(C)=-0{\cdot}1$

(*d*) $P(A)=1{\cdot}\ 1,\ P(B)=0{\cdot}5,\ P(C)=0{\cdot}3.$

2·22. If P denotes a probability measure show that $P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(B\cap C)-P(C\cap A)+P(A\cap B\cap C)$ and generalize the result.

2·23. Let A be the event that a philosopher taking an evening walk in the museum park in Trivandrum on a Sunday evening, will see a girl in a cotton sari and let B be the event that he will see a girl who has anna nada (walking like a swan in a lake). Let the probabilities for these events be given as $P(A)=0{\cdot}50$, $P(B)=0{\cdot}40$ and $P(A\cap B)=0{\cdot}20$. Interpret the events $A\cup B$, $A\cup\bar{B}$, $\bar{A}\cup B$, $\bar{A}\cup\bar{B}$, $A\cap\bar{B}$, $\bar{A}\cap B$ and $\bar{A}\cap\bar{B}$ and evaluate the corresponding probabilities.

2·24. Consider an experiment of throwing a coin 3 times. How many possible events are there ?

2·4. HOW TO ASSIGN PROBABILITIES TO VARIOUS EVENTS

In the previous section we have seen that the probability of an event is a non-negative number which is less than or equal to one. Also we noticed that with these properties for a probability there are a number of possibilities for the probability of an event A. Here we will discuss further considerations which will enable us to determine the probability of an event in some sense.

2·41. Symmetry. Consideration of symmetry in the outcomes of an experiment is a useful tool for deciding the probability of an event. Consider an experiment of rolling a perfectly symmetric die once. If the die is as nearly perfect and symmetric as possible with respect to the six sides marked 1, 2, 3, 4, 5, 6 then it is justifiable to assign equal probabilities to the occurrence of 1, occurrence of 2, ..., occurrence of six. Here the outcome set

$$S=\{1,2,3,4,5,6\}$$
$$=\{1\}\cup\{2\}\cup\{3\}\cup\ldots\ldots\cup\{6\}$$
$$\therefore\quad 1=P(S)=P\{1\}+P\{2\}+\ldots\ldots+P\{6\}.$$

From symmetry we assume that

$$P\{1\}=P\{2\}=\ldots\ldots=P\{6\}$$
$$=\frac{1}{6}.$$

i.e., The probability of getting any one face (say 6) when a symmetric die is rolled once, is 1/6.

Consider another example of tossing a coin which is as symmetric and as perfect as possible with respect to the two sides head and tail. In such a case the coin may be called unbiased. If the possibilities in a trial are head and tail then each can be given a probability 1/2.

Ex. 2.41.1. *If an unbiased coin is thrown 3 times find the probability of getting atleast one head ?*

Solution. The outcomes may be denoted by

$$V_1=(H, H, H) \qquad V_2=(H, H, T)$$
$$V_3=(H, T, H) \qquad V_4=(H, T, T)$$
$$V_5=(T, T, T) \qquad V_6=(T, T, H)$$
$$V_7=(T, H, T) \qquad V_8=(T, H, H)$$

then the event A of getting atleast one head is

$A=\{V_1, V_2, V_3, V_4, V_6, V_7, V_8\}$ and $P(A)=P\{V_1\}+P\{V_2\}+P\{V_3\}+P\{V_4\}+P\{V_6\}+P\{V_7\}+P\{V_8\}$, (since $V_1, V_2, \ldots V_8$ are mutually exclusive elementary events). We may assume equal probabilities 1/8 each for all these elementary events since there is complete symmetry in the experimental outcomes.

Therefore, $P(A)=7/8$.

Comments. From this example it may be noticed that $P(A)=1-P(\bar{A})$ and $\bar{A}$ has only one element V_5 and therefore $P(A)=1-P\{V_5\}=1-1/8=7/8$. This may also be considered to be the total number of elementary events or outcomes or sample points favourable to the event A, divided by the total number of elementary events, that is,

$$P(A)=\frac{\text{Total number of elementary events favourable to A}}{\text{Total number of elementary events}}$$

where there is symmetry in the outcomes. Whenever there is symmetry in the experimental outcomes the above ratio can be used as a tool for assigning probabilities to events.

Ex. 2.41.2. *Find out the probability of getting an ace from a well shuffled deck of 52 card if a card is selected at random.*

Solution. We may assume symmetry because the cards are well shuffled. The total number of elements in the outcome set is 52 and the number of elements favourable to the event is 4 and hence the required probability is 4/52.

Comments. Here at random means that no importance is given to anyone card when selecting a card or all the cards are given equal chances of being selected in the sense that symmetry is not lost while selecting a car .

Ex. 2.41.3. *A consignment of 100 radios contains 25 defective ones. What is the probability of getting 3 defectives in a random sample of size 10 from the consignment ?*

Solution. Here the sample is a random sample *i.e.*, all the samples of size 10 are given equal chances of being selected. We can assume symmetry. The total number of possible samples of size 10 is $\binom{100}{10}$. The number of ways in which a random sample of size 10 containing 3 defectives may be selected

$$=\binom{25}{3}\binom{75}{7}$$

Hence the required probability

$$=\frac{\binom{25}{3}\binom{75}{7}}{\binom{100}{10}}$$

Comments. It is seen that consideration of symmetry is a useful tool for assigning probabilities. But when the outcome set has an infinite number of elements the method of taking the ratio (the number of elements favourable to an event to the total number of elements) must be modified to suit the experiment under consideration.

2.42. Method of Relative Frequencies. Consideration of symmetry may not be suitable for all problems. For example if we want to find out, the probability of getting a head in an experiment of throwing a biased coin, the probability that a new born baby in a particular set of people, is a boy, the probability that a man will die, the probability that a monkey will type this book word by word if it is given a typewriter to play with etc., symmetry is of very little use in finding out the probabilities. It is quite unlikely, but not impossible, that a monkey will type this book word by word if it is given a typewriter to play with. We are quite justified in assigning a probability zero to this event. This is almost surely an impossible event. So far there is no known case of a man living forever. We are justified in assigning a probability one to the event that a man will die. This is almost surely a sure event. *These types of arguments have led to a definition of probability as a measure of conviction of mind based on experience.* In an experiment of throwing a coin if everything is known about the coin such as all physical characteristics of the coin, the forces acting on the coin etc., one may be able to say whether the outcome is head or tail. So some people may argue that probability is in some sense a measure of our ignorance about the various aspects of the experiment. For an elaborate discussion of personal probability, utility etc., the reader is advised to see 'Foundations of Statistics' by L.J. Savage.

Even if there is symmetry, in many cases, a man with a logical mind sometimes gets confused in deciding symmetry in an experiment. For example, consider the following problem. A businessman wants to go for a business trip. Two of his young secretaries Miss Chick and Miss Cute want to accompany him. He wants only one secretary for the trip. He decided to conduct a game of chance and take a decision. He throws an unbiased coin twice. If there is atleast one head he will take Miss Chick. Otherwise Miss Cute will be taken. What is the probability that Miss Chick will be taken ? This may be argued like this. There are four possibilities (H, H), (H, T), (T, H) and (T, T) ; out of these 3 are favourable to Miss Chick's selection. So the probability is 3/4. Some people may argue like this. If head comes in the first trial the experiment is over and Miss Chick is selected. Therefore the possible outcomes are (T, H), (T, T) and H, and hence the required probability is 2/3.

These show that consideration of symmetry alone does not uniquely determine the probability of an event. Another method of determining the probability of an event is consideration of relative frequencies. Suppose that we want to find out the probability of getting a head in throwing a coin, assuming that the coin will fall either head or tail, but nothing is known about the biasedness of the coin. We can estimate the probability of getting a head by conducting an experiment. Throw the coin under the same conditions without giving any importance to any side 100 times. Count the number of heads. Take the relative frequency —the ratio of the number of heads to the total number of trials. Repeat the experiment 1000 times and take the relative frequency. Continue the experiment. If this relative frequency tends to a limit in the long run, this limit may be taken as the probability of getting a head if that coin is thrown under the conditions of the experiment. If the experiment is conducted a sufficiently large number of times then a good estimate of the probability is given by the relative frequency. There are other questions in this respect. We do not know whether there exists a limit or not. If a limit does not exist our estimate of the probability has no meaning. Further the repeatability of an experiment under the same conditions, is assumed.

The discussion of the definition and the methods of evaluation of probability is not complete in this section. For further reading see the references at the end of this chapter.

2·5. SOME USEFUL CORRESPONDENCE BETWEEN SET THEORY AND PROBABILITY THEORY

	Set Theory	*Probability Theory*
1.	Outcome set S. (sample space, possibilities space etc.)	Sure event.

Set Theory	Probability Theory
2. Subset of an outcome set	Event.
3. Element of the outcome set	Elementary event.
4. Disjoint subsets.	Mutually exclusive events.
5. Null set ϕ.	Impossible event.
6. Totally additive, non-negative, set function with total measure unity.	Probability measure.
7. P(A) = 0	A is almost surely an impossible event.
8. P(A) = 1	A is an almost sure event.
9. $A \cup B$ (where A and B are subsets of S).	Atleast one of the events A and B.
10. $A \cap B$.	Simultaneous occurrence of A and B.
11. $\bar{A}$ (Complement of A).	Non-occurrence of the event A.

Exercises

2·25. In a class there are 30 boys and 20 girls. From the class list one name is picked up at random. What is the probability that it is a boy's name ?

2·26. A bridge player has 7 spades. What is the probability that his partner has

(*a*) 2 spades ?

(*b*) at least 2 spades ?

2·27. What is the probability of throwing 7, 8 or 9 with 2 dice ?

2·28. In a set of 100 businessmen 10 are unmarried and others are married. What is the probability of selecting a sample of 20 businessmen out of which 15 are married, if the sample is selected at random ?

2·29. In a community of 400 people 200 are highly intelligent, 100 are above average, 50 are average and the rest are idiots. If a random sample of size 40 is taken, what is the probability that the sample contains 20 highly intelligent, 10 above average and 10 average ?

2·30. A psychiatrist reports that his survey of 100 people on a lunar ecclipse shows that 70 are psychopaths, 50 have xenophobia, 20 are psychopaths and have xenophobia and 10 are neither psychopaths nor have xenophobia. Should you question the results in his findings ?

2.6. CONDITIONAL PROBABILITY

In the theory discussed so far we were concerned about probabilities of events (subsets of an outcome set S) relative to an outcome set S. In order to speak of probabilities of events, an outcome set is to be specified.

P(A) is the probability measure of A relative to S.. This may be denoted, for convenience, as P(A | S) (probability of an event A relative to the set S or probability of A given S) So a probability statement such as "The probability of an event A is 0·95", is a conditional statement or a relative statement relative to a specified outcome set S. In the following discussion we will consider

probabilities of events relative to various events, for example probability statements like $P(A \mid B)=0{\cdot}3$ (probability of the event A given the event B is 0·3). When B is the outcome set S we will denote the conditional probability $P(A \mid S)$ as $P(A)$, otherwise we will denote the conditional probability by $P(A \mid B)$.

Ex. 2.6.1. *Consider the following problem. In Dr. Nimbus' reducing laboratory people are given one of the two treatments, VL 50 and VL 100. There are 80 people taking treatment, out of which some are divorced and looking for the next husband and some are unmarried, some weigh more than 150 lbs. and some weigh less than 150 lbs. The following is the exact classification.*

		Over 150 lbs.	*Less than 150 lbs.*
S_1	*Divorced*	*15*	*10*
	Unmarried	*20*	*5*

VL 50

		Over 150 lbs.	*Less than 150 lbs.*
S_2	*Divorced*	*12*	*1*
	Unmarried	*15*	*2*

VL 100

A visitor Miss Iceberg goes to Dr. Nimbus' laboratory. Assuming that every patient in the laboratory is given an equal chance of being called to give testimony of her achievement in weight reduction, what is the probability that Miss Iceberg will hear testimony from a patient who is (1) undergoing treatment VL 50, (2) divorced, (3) heavier than 150 lbs. ?

Solution. Here there are 80 patients and all are assumed to have an equal chance of being called. For the probabilities in (1), (2) and (3) we are concerned with the outcome set $S=S_1 \cup S_2$ (set of all the patients). If the above events are denoted by A, B, C respectively then

$$P(A)=\frac{(15+10+20+5)}{80}$$

$$=\frac{50}{80}$$

$$P(B)=\frac{(15+10+12+1)}{80}$$

$$=\frac{38}{80}$$

$$P(C)=\frac{(15+20+12+15)}{80}$$

$$=\frac{62}{80}.$$

What is the probability that Miss Iceberg will hear a person who is over 150 lbs given that the speaker is divorced ? This can be denoted by $P(C \mid B)$=probability of the event C given B. Here we are concerned only with the divorced patients. So the set under consideration is S_3 the set of divorced persons. Where S_3 is given by,

		Over 150 lbs.	Less than 150 lbs.
S_3	VL 50	15	10
	VL 100	12	1

Total=38

Hence $$P(C \mid B)=\frac{(15+12)}{38}$$

$$=\frac{27}{38}.$$

Comments. In this example it may be noticed that $P(C\cap B)$=probability that the speaker is divorced and over 150 lbs. in weight=27/80. It may be further noticed that

$$\frac{P(C\cap B)}{P(B)}=\frac{27}{80}\div\frac{38}{80}$$

$$=\frac{27}{38}=P(C \mid B).$$

Further in the reduced sample space or outcome set of divorced persons the probabilities of the mutually exclusive events of getting a person, who weighs more than 150 lbs. and is taking VL 50, who weighs less than 150 lbs. and is taking VL 50, who weighs more than 150 lbs. and is taking VL 100, who weighs less

than 150 lbs. and is taking VL 100, add up to unity. That is probabilities of the form $P(A \mid B)$ add up to unity in the reduced set B. With these intuitive notions we will define conditional probability as follows.

2·61. Definition. If A and B are events and if $P(B) \neq 0$ then the conditional probability of A given B may be defined as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

i.e., The probability of A in the reduced set B equals the probability of $A \cap B$ in the outcome set S, divided by the probability of B is the outcome set S. From the above definition we have a general multiplication rule as follows.

$$P(A \cap B) = P(A \mid B)\, P(B)$$
$$= P(B \mid A)\, P(A).$$

This rule can be extended to a number of events. For example

$$P(A \cap B \cap C) = P(A \mid B \cap C) \,.\, P(B \cap C)$$
$$= P(A \mid B \cap C) \,.\, P(B \mid C) \,.\, P(C) \text{ etc.}$$

Ex. 2.61.1. *A consignment of 20 radios contains 6 defectives. Radios are selected at random one by one and examined. The radios examined are not put back. What is the probability that the 10th one examined is the last defective?*

Solution. There are exactly 5 defectives in the first 9 examinations. Let A denote the event of getting 5 defectives in the first 9 examinations.

$$P(A) = \frac{\binom{14}{4}\binom{6}{5}}{\binom{20}{9}}$$

Let B be the event of getting the 10^{th} examined a defective

$$P(B \mid A) = \frac{1}{11}.$$

(There are 11 radios left, out of which one is a defective).

$\therefore$ The probability that the 10^{th} one is the 6^{th} defective

$$= P(A \cap B) = P(A) \,.\, P(B \mid A)$$

$$= \frac{1}{11} \cdot \frac{\binom{14}{4}\binom{6}{5}}{\binom{20}{9}}$$

$$= \frac{21}{6460}$$

Ex. 2.61.2. *A box contains 3 red balls and 7 green balls. 2 balls are picked out one by one at random without replacement. What is the probability that the second ball is green given that the first one is green?*

Solution. If the first one is green and if it is not put back then there are 6 green balls out of 9 balls and hence the required probability is 6/9.

2.62. Independence of Events. Two events A and B are said to be independent if $P(A \cap B) = P(A).P(B)$. Independent events may be explained as events where the occurrence of one does not affect the occurrence or non-occurrence of the other. This explanation is often misleading so the reader is advised to stick to the definition as $P(A \cap B) = P(A).P(B)$.

Ex. 2·62·1. *In example 2·61·2 what is the probability of drawing two green balls if the first ball is replaced.*

Solution. Here the probability of drawing the first green ball is 7/10. Second trial is independent of the first trial and the probability of getting a green ball is 7/10. Hence the required probability is $\frac{7}{10} \cdot \frac{7}{10} = 49/100$.

Theorem. 2.62.1. If A and B are independent events and if $P(B) \neq 0$ then, $P(A \mid B) = P(A)$

Proof. If A and B are independent $P(A \cap B) = P(A)\ P(B)$

$$\text{But } P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A).P(B)}{P(B)} = P(A).$$

Theorem. 2.62.2. If A and B are independent prove that (1) $\bar{A}$ and $\bar{B}$ are independent, (2) $\bar{A}$ and B are independent, (3) A and $\bar{B}$ are independent.

i.e, (1) Prove that $P(\bar{A} \cap \bar{B}) = P(\bar{A}).P(\bar{B})$ given that $P(A \cap B) = P(A).P(B)$

Proof.
$$\begin{aligned} P(\bar{A} \cap \bar{B}) &= P(\overline{A \cup B}) = 1 - P(A \cup B) \\ &= 1 - [P(A) + P(B) - P(A \cap B)] \\ &= 1 - P(A) - P(B) + P(A).P(B) \\ &= [1 - P(A)][1 - P(B)] = P(\bar{A}).P(\bar{B}) \end{aligned}$$

(2) and (3) are left to the reader.

2.62. Pairwise independence. The events $A_1, A_2, \ldots, A_n$ are said to be pairwise independent if $P(A_i \cap A_j) = P(A_i).P(A_j)$ for all i and j, $i \neq j$.

2.63. Mutual independence. The events $A_1, A_2, \ldots, A_n$ are mutually independent if the probabilities of the intersection of any

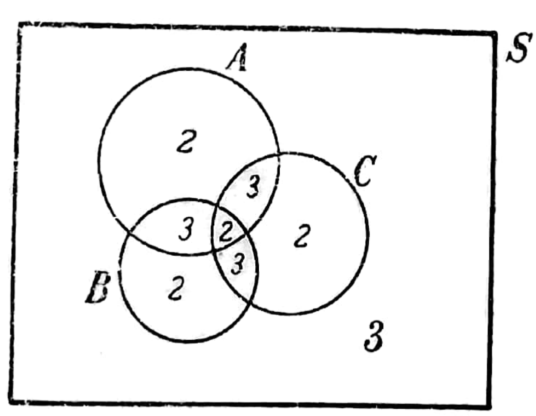


Fig. 2·13.

two, any three,...any n events are the products of the respective individual probabilities. Pairwise independence does not mean mutual independence. As an illustration consider the following Venn diagram where $P(A)=P(B)=P(C)=1/2$ $P(A\cap B)=P(B\cap C)$ $=P(C\cap A)=1/4$ and $P(A\cap B\cap C=1/10$.

Here the numbers in the various subsets denote the numbers of elementary events in the various events and symmetry in the outcomes is assumed.

Lemma. 2.62.3. If $B_1, B_2,...B_n$ denote a disjoint partition of the outcome set and if $P(B_i)\neq 0$ for $i=1, 2,...,n$ then for any

event A, $P(A) = \sum_{i=1}^{n} P(B_i).P(A/B_i)$

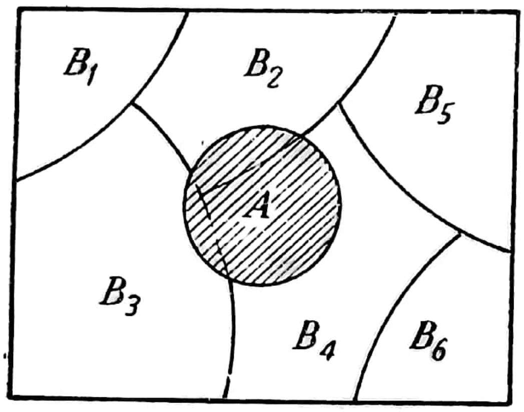


Fig. 2.14.

Proof. In this case A may be written as

$$A=(A\cap B_1)\cup(A\cap B_2)\cup\cdots\cup(A\cap B_n).$$

(This may be verified from Fig. 2·14.)

where $A\cap B_1$, $A\cap B_2,...,A\cap B_n$ are all disjoint

$$\therefore \quad P(A)=P(A\cap B_1)+P(A\cap B_2)+\cdots+P(A\cap B_n)$$

$$=P(B_1)\,P(A\mid B_1)+P(B_2)P(A\mid B_2)+\cdots+P(B_n)\times P(A\mid B_n)$$

$$=\sum_{i=1}^{n} P(B_i)\,P(A\mid B_i).$$

Ex. 2.62.2. *Four people Mr. Fox, Miss Tod, Mr. Rock and Mr. Jack compete for the presidency of Piggyland. A public opinion poll gives the estimates of the probabilities of their winning as 0·40, 0·20, 0·30 and 0·10 respectively. The probability that gambling will be nationalized by them, if they are elected are 0·85, 0·90, 0·30 and 0·95 respectively.*

What is the probability that gambling will be nationalized after the presidential election ?

Solution. Let A be the event that gambling will be nationalized after the presidential election.

Let B_1, B_2, B_3 and B_4 be the events that Mr. Fox, Miss Tod, Mr. Rock and Mr. Jack will be elected, respectively. Then we are given that

$$P(B_1)=0{\cdot}40,\ P(B_2)=0{\cdot}20,\ P(B_3)=0{\cdot}30 \text{ and } P(B_4)=0{\cdot}20$$

$P(A \mid B_1)=0{\cdot}85$, $P(A \mid B_2)=0{\cdot}90$, $P(A \mid B_3)=0{\cdot}30$
and $P(A \mid B_4)=0{\cdot}95$

$$\therefore\ P(A)=\sum_{i=1}^{4} P(B_i)\,P(A \mid B_i)$$

$$=0{\cdot}40\times0{\cdot}85+0{\cdot}20\times0{\cdot}90+0{\cdot}30\times0{\cdot}30+0{\cdot}20\times0{\cdot}95$$
$$=0{\cdot}80.$$

Theorem 2.62.4. Bayes' Rule. If $B_1, B_2, \ldots B_n$ denote a disjoint partition of the outcome set, $P(B_i)\neq0$ for $i=1, 2\ldots n$ and $P(A)\neq0$ then

$$P(B_r \mid A)=\frac{P(B_r)\,.\,P(A \mid B_r)}{\sum_{i=1}^{n} P(B_i)\,.\,P(A \mid B_i)}$$

for $r=1, 2,\ldots n$.

Proof. By definition

$$P(B_r \mid A)=\frac{P(B_r\cap A)}{P(A)}=\frac{P(B_r).P(A \mid B_r)}{P(A)}$$

By lemma 2·62·3. $P(A)=\sum_{i=1}^{n} P(B_i)\,P(A \mid B_i)$.

Hence the result.

Here in this theorem the $P(B_i)$ for $i=1, 2,\ldots n$ may be called prior or a priori probabilities in the sense that they are probabilities determined before the observation of the occurrence of any event and $P(B_i \mid A)$ for $i=1, 2,\ldots n$ may be called posterior or a posteriori probabilities in the sense that these are probabilities of B_i for $i=1, 2,\ldots n$ after observing A. This theorem is aften called Bayes' Rule and is important because it gives some sort of inverse reasoning. This aspect may be seen from the following example. Bayes assumed the principle of equal division of ignorance in an unknown situation, *i.e.*, he assumed that all the prior probabilities $P(B_i)$ for $i=1, 2,\ldots n$ are equal if nothing is known about them. There is a lot of controversy regarding this point. For further information see the references given at the end of this chapter.

Ex. 2.62.3. *Three machines X, Y, and Z of equal capacities are producing bullets. The probabilities that the machines produce defectives (bullets which do not satisfy the specifications) are 0·1, 0·2 and 0·1 respectively. A bullet is taken at random from a day's production and is found to be defective. What is the probability that it came from machine X ?*

Solution. Let B_1, B_2, B_3 be the events of getting a bullet which came from the machines X, Y, Z respectively. Let A be the event of getting a defective bullet. Since the machines are of equal capacity we can assume that, $P(B_1)=P(B_2)=P(B_3)=1/3$.

$P(A \mid B_1)=$probability of getting a defective given that it came from $X=0{\cdot}1$.

$$P(A \mid B_2)=0\cdot 2 \text{ and } P(A \mid B_3)=0\cdot 1.$$

We want the probability that the defective came from $X=P(B_1 \mid A)=$Probability of getting the bullets which came from X given that it is defective and, therefore,

$$P(B_1 \mid A)=\frac{P(B_1)\, P(A \mid B_1)}{\Sigma^3_{i=1}\, P(B_i)\, P(A \mid B_i)}$$

$$=\frac{(1/3)(0\cdot 1)}{(1/3)(0\cdot 1)+(1/3)(0\cdot 2)+(1/3)(0\cdot 1)}$$

$$=0\cdot 1/0\cdot 4=1/4$$

Exercises

2.31. Find the probability of drawing two spades from a well shuffled deck of 52 cards if (1) the first card is replaced before the second one is taken, (2) the first one is not replaced.

2.32. Give one example each of two events which are (*a*) mutually exclusive and independent, (*b*) mutually exclusive but not independent (*c*) not mutually exclusive but independent, (*d*) not mutually exclusive and not independent.

2.33. Give 2 examples of three events which are pairwise independent but not mutually independent.

2.34. A radio station broadcasts the correct time every hour on the hour. What is the probability that a listner who switches on the radio at random has to wait less than 20 minutes to hear the correct time ?

2.35. A line cuts the line segment A B into two parts. What is the probability that the ratio of the two segments (smaller to the larger) is less than 1/4 ?

2.36. If A and B are mutually exclusive events and if $P(A)=0\cdot 5$ and $P(B)=0\cdot 3$ find the probability of (*a*) $A \cup B$, (*b*) $A \mid B$, (*c*) $\bar{A} \cap \bar{B}$.

2.37. There are three machines producing 10,000, 20,000, 30,000 bullets per hour respectively. These machines are known to be producing 1%, 2%, 1% defectives respectively. One bullet is taken at random from an hour's production of the three machines and found to be defective. What is the probability that this bullet came from the third machine ?

2.38. From a box containing 5 red and 10 white rose flowers two flowers are taken at random one by one. If the first one is a red rose, what is the probability that (*a*) the second one is red, (*b*) the second one is white ?

2.39. A survey is conducted on two random samples of 100 men and 100 women ; it is seen that 2 men are deaf and one woman is deaf. Taking these proportions as estimates of the probabilities of getting a deaf man and a deaf woman respectively, what is the probability that a deaf person taken at random is a male ?

2.40. Among three identical urns one has 2 red marbles, one has one red and one green marble and the third has 2 green marbles. One urn is selected at random and then a marble is picked at random. It is found to be red. What is the probability that the other marble in the urn is also red ?

2.41. Suppose that one of three men, a politician, a businessman an educator will be appointed as the chancellor of a university. The respective probabilities of their appointments are 0·50, 0·30, 0·20. The probabilities that research activities will be promoted by these people if they are appointed are 0.30, 0·70, 0·80 respectively. What is the probability that research will be promoted by the new chancellor ?

2.42. There is a chance for a particular engineering project's failure whether machine X fails or not. The probability that machine X fails is 0·1. The probability of the project's failure if X fails is 0·9 and is otherwise 0·2. It is seen that the project has failed. What is the probability that the failure is due to the failure of X ?

2·7. ENTROPY OF A FINITE SCHEME

This concept is widely used in Information Theory and in the Theory of Communications. We will give a brief introduction in this section. For additional reading in this line references are given at the end of this chapter.

2·71. A Complete System of Events. A system of events, $A_1, A_2, \ldots, A_n$ in which one and only one of them occurs in each trial, may be called a complete system of events.

For example,

(*a*) Appearance of head or tail in the throwing of a coin.

(*b*) Appearance of 1 or 2 or 3 or 4 or 5 or 6 in the rolling of a die once.

(*e*) A set of mutually exclusive events $A_1, A_2, \ldots\ldots A_n$ such that $A_1 \cup A_2 \cup \ldots\ldots \cup A_n = S$ (the outcome set).

2·72. Finite Scheme. A complete system of events $A_1, A_2, \ldots A_n$ where n is finite, together with their probabilities $P_1, P_2, \ldots\ldots P_n$ is called a finite scheme. By this definition a finite scheme may be represented by the matrix

$$A = \begin{bmatrix} A_1, & A_2, \ldots\ldots A_n \\ P_1, & P_2, \ldots\ldots P_n \end{bmatrix}$$

when $A_1 \cup A_2 \cup \ldots\ldots \cup A_n = S$ and $A_i \cap A_j = \phi$ for all i and j, $i \neq j$

$0 \leqslant P_i \leqslant 1$ and $\sum_{i=1}^{n} P_i = 1$.

2·73. Entropy of a Finite Scheme. Let us consider three finite schemes

$$(a) \quad \begin{pmatrix} A_1, & A_2 \\ 0{\cdot}5, & 0{\cdot}5 \end{pmatrix}, \; (b) \begin{pmatrix} A_1, & A_2 \\ 0{\cdot}999, & 0{\cdot}001 \end{pmatrix}, \; (c) \begin{pmatrix} A_1, & A_2 \\ 0{\cdot}4, & 0{\cdot}6 \end{pmatrix}.$$

In (*a*) A_1 and A_2 have equal chances of occurrence and therefore there is great uncertainty about the occurrence of A_1 or A_2. In (*b*) there is a very great chance of A_1's occurrence. So in this scheme there is much less lack of certainty. In (*c*) the lack of certainty may be said to be in between those in (*a*) and (*b*). It is desirable to have a measure of lack of certainty in a finite scheme. One of the measures suggested is $H(p_1, p_2, \ldots p_n) = -k \sum_{i=1}^{n} p_i \log p_i$ where k is a positive constant and H is only a notation for a function of $p_1, \ldots\ldots, p_n$, and $\log p_i$ means the natural logarithm of p_i (if $a^x = b$ then x is called the logarithm of b to the base a and is written as $x = \log_a b$. When $a = e$ the base is not usually written, *i.e.*, $e^x = b$ is written as $x = \log b$. These are called natural logarithms where e is a constant approximately equal to 2·71818). $H(p_1, \ldots, p_n)$

is usually called the entropy of the finite scheme $\begin{pmatrix} A_1, \ldots, & A_n \\ p_1, \ldots, & p_n \end{pmatrix}$. In practical situations, $H = -\Sigma\ p_i \log p_i$ (*i.e.*, $K=1$) serves the purpose and so for convenience H in this simplified form will be taken here. It is also called the 'information' in the finite scheme. This definition of 'information' is used in 'Information Theory' and in 'Mathematical Theory of Communications' etc. Some statistical concepts of 'information' will be discussed in later chapters. Some other areas where the probability theory is widely applied are (1) Markov Processes, (2) Ergodic Theory, (3) Random Walk, (4) Queuing Theory, (5) Genetics, (6) Space Research—especially in predicting the operational ability of space vehicles, region of impact of bombs, rockets etc. (7) Agricultural production, in testing of crop yields, and in selection of one variety from the other, (8) Industrial production process, especially in controlling the quality of goods, (9) Sociological, Psychological and Public opinion surveys etc.

Exercises

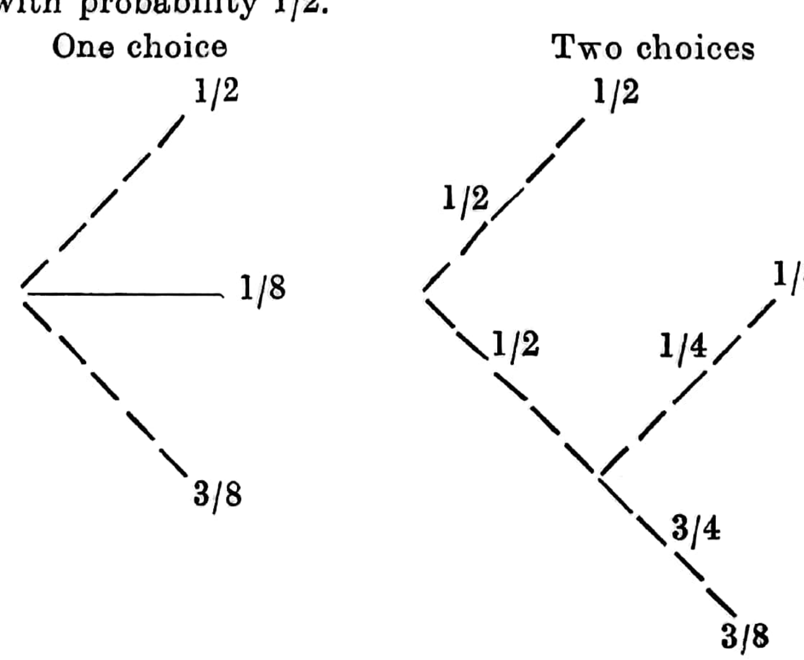
2·43. Find the information $H(p_1, \ldots\ldots, p_n) = -\Sigma\ p_i \log p_i$ in the finite scheme,

$$\begin{pmatrix} A_1, & A_2, & A_3, & A_4 \\ \cdot 2, & \cdot 4, & \cdot 3, & \cdot 1 \end{pmatrix}$$

2·44. In a discrete noiseless system (a message is not disturbed while it travels) a coded message source produces a sequence of letters chosen from among the letters *a, b, c, d* with probabilities 1/10, 1/5, 2/5, 3/10 respectively, where successive symbols are chosen independently. What is the entropy per symbol ? Show that it is a maximum when the probabilities are equal. (This is intuitively the most uncertain situation).

2·45. Show that the entropy $H(p_1, \ldots\ldots, p_n)$ in a finite scheme satisfies the following conditions.

(1) H is continuous in p_i ; (2) If all the p_i's are equal, that is, if $p_i = 1/n$ for all i, then H is a monotonic increasing function of n. (This implies that with equally likely events there is more choice or more uncertainty) ; (3) If a choice consists of two successive choices the original H is a weighted sum of the individual values of H. For example if the choices are shown in the tree diagram below, then $H(1/2,\ 1/8,\ 3/8) = H(1/2,\ 1/2) + (1/2)H(1/4,\ 3/4)$ where the weight 1/2 is taken because the second choice occurs with probability 1/2.



It can be shown that any function satisfying the conditions (1), (2) and (3) is of the form $-k \; \Sigma p_i \log p_i$ where k is a positive constant.

BIBLIOGRAPHY

1. H. **Cramer.** "Elements of Probability Theory" John Wiley and Sons, N.Y. 1955.

2. H. **Cramer,** "Mathematical Methods of Statistics" Princeton University Press, N.J. 1961.

3. W. **Feller,** "Probability Theory and its Applications" Vol. I. John Wiley and Sons, N.Y. 1950.

4. D.A.S. **Eraser,** "Statistics, An Introduction" John Wiley and Sons, N.Y. 1958.

5. J.E. **Freund**. "Mathematical Statistics" Prentice-Hall, N.J. 1954.

6. H. **Jeffreys**, "Theory of Probability," Clarenden Press, Oxford, 1939.

7. D.V. **Lindley,** "Introduction to Probability and Statistics" Cambridge University Press, 1965.

8. M. **Loeve,** "Probability Theory" Van Nostrand, Princeton, N.J., 1960,

9. M.E. **Munroe,** "Theory of Probability. McGraw Hill, N.Y., 1951.

10. M.E. **Munroe,** "Introduction to Measure and Integration", Addison-Wesley, Reading, Mass, 1959.

11. E. **Parson,** "Modern Probability Theory and its Applications" John Wiley and Sons, N.Y., 1960.

12. M. **Richardson,** C. "College Algebra" Prentice-Hall, Inc., N.J., 1965.

13. L.J. **Savage,** "Foundations of Statistics" John Wiley and Sons, N.Y., 1954.

14. C.E. **Shannon** and W. **Weaver,** "Mathematical Theory of Communication" The University of Illinois Press, Urbana 1963.

15. J. **Uspensky,** "Introduction to Mathematical Probability" McGraw Hill, 1937.

CHAPTER 3

STOCHASTIC VARIABLES

Introduction. It is assumed that the reader is already familiar with mathematical variables, differential and integral operators etc. In this chapter we will introduce another type of variable called a stochastic variable, an operator called mathematical expectation etc. In this chapter only the one variable or one variate case is considered.

Stochastic Variables. An outcome set was seen to be finite or infinite, discrete in the sense individually distinct or continuous, real or hypothetical. It was also noticed in chapter 2 that probability is a measure defined on the outcome set S such that the total measure $P(S)=1$. Now we will consider other types of functions defined on the outcome set. One such function is called a stochastic variable. A Stochastic variable X is a real valued function defined on the outcome set S such that its domain is the outcome set and its range is the real line $R=(-\infty, \infty)$. (Here R means the set of all real numbers from $-\infty$ to $+\infty$). Evidently these can be represented on a real line). In this book we will deal only with real stochastic variables. *i.e.*, functions whose range is the real line or the set of real numbers. If a stochastic variable is defined on a discrete outcome set it is called a discrete stochastic variable and if it defined on a continuous outcome set then it is called a continuous stochastic variable. Stochastic variables are also called chance variables, random variables, variates etc. For a rigorous definition, see reference [3] at the end of this chapter.

3.11.1. DISCRETE STOCHASTIC VARIABLES

Consider the following experiment of throwing a coin twice. If the occurrence of a head is denoted by 1 and that of a tail by 0, the outcome set, which is discrete, consists of the outcomes $(0, 0), (0, 1), (1, 0), (1, 1)$. *Let X denote the total number of heads in an outcome of the experiment. Evidently X is a discrete stochastic variable defined on* $S=\{(0, 0), (0, 1), (1, 0), (1, 1)\}$. *There can only be 0 or 1 or 2 heads. So the range of X is the set* $A=\{0, 1, 2\}$ *and evidently this set A is contained in the real line* $R=(-\infty, \infty\}$. *For the same experiment let Y denote the quantity "number of heads minus number of tails" in the outcomes of this experiment. Evidently Y is a discrete stochastic variable with the range* $B=\{-2, 0, 2\}$.

Comments. Stochastic variables are usually denoted by X, Y, Z etc., and their ranges by x, y, z etc. For example in Ex. 3.11.1., any element in the range of X may be denoted by x, then x takes the values 0, 1 and 2. It may be noticed that a number of stochastic variables may be defined on a given outcome set.

3.11.2. *Consider the experiment of rolling a die twice. Let X denote the sum rolled. (that is the sum of the two face numbers). Evidently X is a stochastic variable which assumes the values 2, 3, 4, 5,..., 12.*

Comments. In Ex. 3.11.1 and 3.11.2 it may be noticed that we can always attach a probability corresponding to every value a discrete stochastic variable takes. (Hereafter we shall use the abbreviation *s.v.* for a stochastic variable). In Ex. 3.11.1, X takes the values 0, 1, 2 with probabilities 1/4, 2/4, 1/4 respectively; Y takes the values −2, 0, 2 with probabilities, 1/4, 2/4, 1/4 respectively. In Ex. 3.11.2, X takes the values 2, 3, 4,..., 11, 12 with probabilities 1/36, 2/36,..., 1/36 respectively. This probability is called the probability function of a stochastic variable X and is usually denoted by $f(x)$. In Ex. 3.11.1 probability functions for X and Y may be defined as follows :

For the *s.v.*X,

$$\begin{array}{c|l} x & f(x) \\ \hline 0 & f(0)=1/4 \\ 1 & f(1)=2/4 \\ 2 & f(2)=1/4 \end{array} \qquad \text{or } f(x)=\begin{cases} 1/4 \text{ for } x=0 \\ 2/4 \text{ for } x=1 \\ 1/4 \text{ for } x=2 \, ; \end{cases}$$

for *s.v.*Y,

$$\begin{array}{c|l} y & f(y) \\ \hline -2 & f(-2)=1/4 \\ 0 & f(0)=2/4 \\ 2 & f(2)=1/4 \end{array} \qquad \text{or } f(y)=\begin{cases} 1/4 \text{ for } y=-2 \\ 2/4 \text{ for } y=0 \\ 1/4 \text{ for } y=2 \, ; \end{cases}$$

for *s.v.X.* in Ex. 3.11.2,

$$\begin{array}{c|l} x & f(x \\ \hline 2 & f(2)=1/36 \\ 3 & f(3)=2/36 \\ \vdots & \\ 12 & f(12)=1/36 \end{array} \qquad \text{or } f(x)=\begin{cases} 1/36 \text{ for } x=2 \\ 2/36 \text{ for } x=3 \\ \quad\vdots \\ 1/36 \text{ for } x=12. \end{cases}$$

As the range of any *s.v X*, according to the definition, is the real line we can define the probability function for X for all the points on the real line. For example consider the *s.v.*'s X and Y in Ex. 3.11.1. The corresponding probability functions may be defined as follows :

$$f(x)=\begin{cases} 1/4 \text{ for } x=0 \\ 2/4 \text{ for } x=1 \\ 1/4 \text{ for } x=2 \\ 0 \text{ elsewhere} \end{cases} \qquad \text{or } f(y)=\begin{cases} 1/4 \text{ for } y=-2 \\ 2/4 \text{ for } y=0 \\ 1/4 \text{ for } y=2 \\ 0 \text{ elsewhere} \end{cases}$$

That is, X assumes the values 0, 1 and 2 with non-zero probabilities as given above and all other values with zero probabilities. Similarly Y assumes the values -2, 0 and 2 with non-zero probabilities and other values with zero probabilities.

Ex. 3.11.3. *Consider a spinner as shown in Fig. 3.1 where the dial is marked from 0 to 100. Suppose that the spinner is completely balanced, in the sense that, the indicator when rotated is as likely to stop at any point as at any other point on the dial.*

If x denotes the distance in the clockwise direction from the point O to the point where the indicator stopped then X is a continuous *s.v.* which can take any value in the interval 0 to 100. Intuitively, the probability that, the indicator when rotated, will stop in between any two points, say, 25 and 32 is $(32-25)/100 = 7/100$. In other words the probability that $25 \leqslant x \leqslant 32$ is $7/100$. To this *s v.* X we can attach a probability function $f(x)$ as follows :

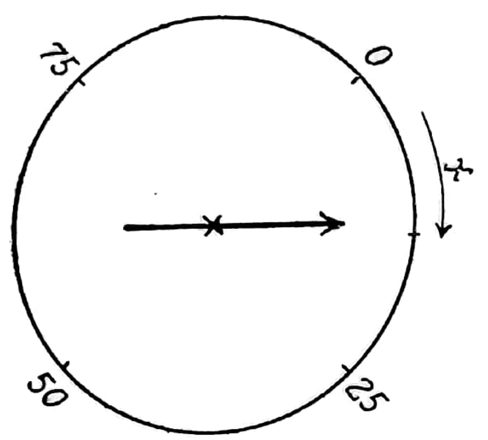


Fig. 3·1.

$$f(x) = \begin{cases} 1/100 \text{ for } 0 \leqslant x \leqslant 100 \\ 0 \text{ elsewhere.} \end{cases}$$

3.12. Graphical Representation. A stochastic variable X and its probability function $f(x)$ may be represented graphically by taking x along one axis, say the X_1-axis and $f(x)$ along the other axis, say the X_2-axis. There are different types of diagrams used frequently for such representations. They are probability curves, bar diagrams, histograms, pie diagrams, pictograms etc. The probability function in Ex. 3.11.3, when represented graphically, gives a curve as shown in Fig. 3.2*a*. In general when the stochastic variable is continuous we can expect the probability function which when represented graphically, to give a continuous curve. One such curve is shown in Fig. 3.2*b*.

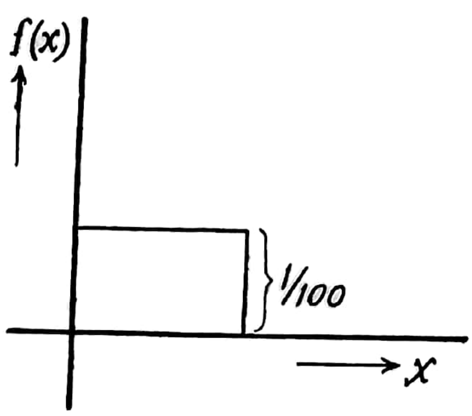


Fig. 3·2(*a*)

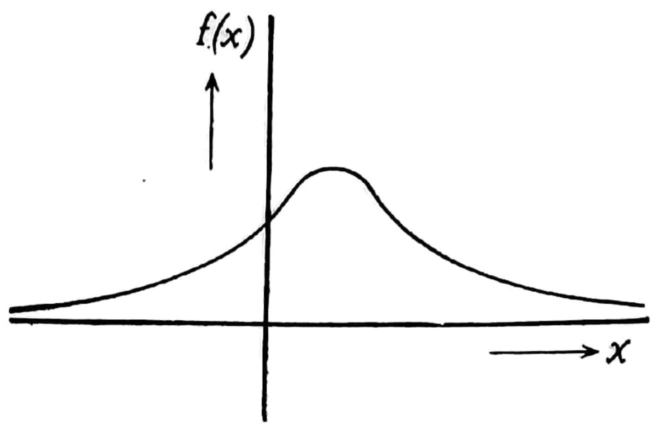


Fig. 3·2(*b*)

The following sections give some of the usual representations of discrete probability functions.

3.13. Bar Diagrams. Fig. 3.3 is a diagrammatic representation of the probability function in Ex. 3.11.2. here bars whose areas are proportional to the probabilities $f(x)$ are erected over the corresponding x. For example in Ex. 3.11.2 the *s.v.* has the range $A=\{2, 3, 4, \ldots, 12\}$ with nonzero probabilities, 1/36, 2/36,..., 1/36 respectively. Thus the area of the bar over the point 2 is propor-

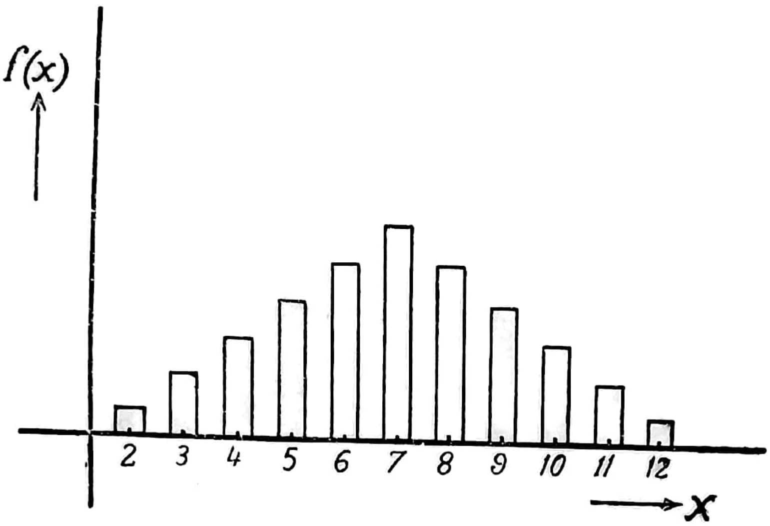


Fig. 3.3.

tional to 1/36, the area of the bar over the point 3 is proportional to 1/36 etc. This bar diagram gives some idea about the probability function to a layman. This technique of bar diagrammatic representation is used for such data as the height measurements of citizens in a city classified according to age groups etc., since relative frequencies estimate probabilities. Here bars proportional to the frequency or number of measurements in a particular age group, may be erected over the middle point of the age interval.

3.14. Histograms. In this diagram the probability function $f(x)$ is represented by rectangles whose areas are proportional to the various probabilities. These rectangles are erected over the intervals such that the points 2, 3,..., 12 are the middle points of the intervals. Here the probability function $f(x)$ in Ex. 3.11.2 is represented by a histogram. This same technique may be used in

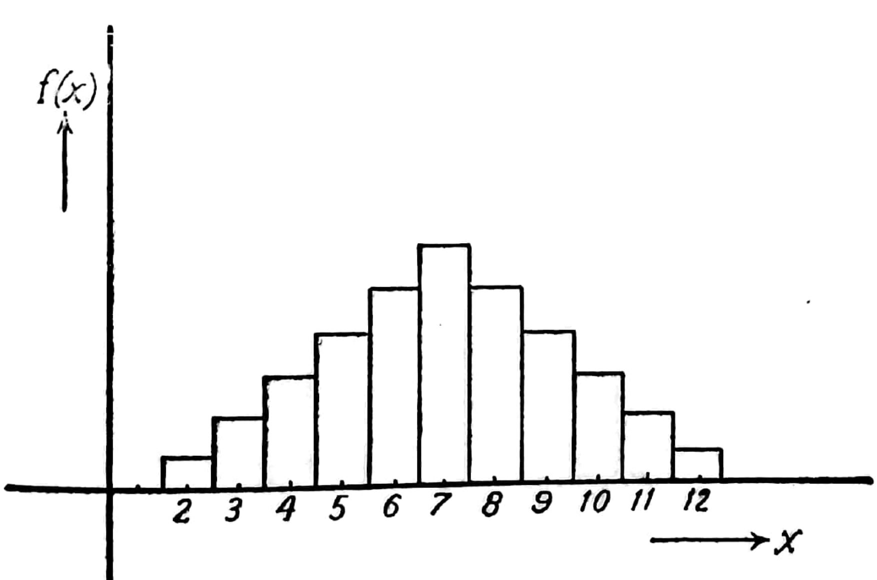


Fig. 3.4.

representing a given classified data in histogram, by erecting rectangles over the various classes such that the areas of the rectangles are proportional to the frequencies in the various classes of a classified data. A representation of discrete probabilities by histograms may give a false impression that x can take all the values in an interval on the real line with nonzero probabilities since our histograms cover an interval. Here we do not assume continuity for x but this representation may help a statistician to get some ideas about the nature of the approximating probability function if a discrete probability function is approximated by a continuous probability function for convenience of mathematical operations.

3.15. Pie Diagrams. If instead of bars or rectangles, the probabilities are represented by sectors in a circle then such a diagrammatic representation is called a pie diagram.

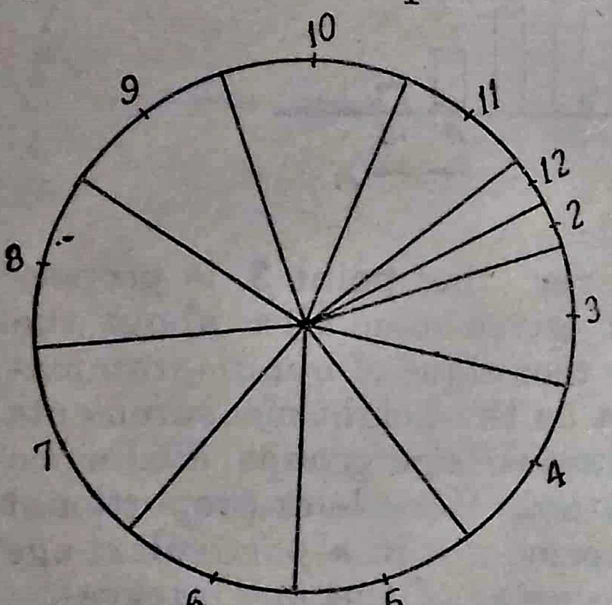


Fig. 3.5.

3.16. Pictograms. Usually data obtained from surveys, experiments, economic data, budgetary data etc., are represented by diagrams so that a non-statistician can have some idea about the allocations of the data into various subdivisions. For example the strength of armies of various countries may be represented by pictures of men, say, one man for every 10,000 men in the army. Such pictorial representations are called pictograms. In Fig. 3.6 the production of apples in three countries is represented in a pictogram to give one some idea about the relative outputs.

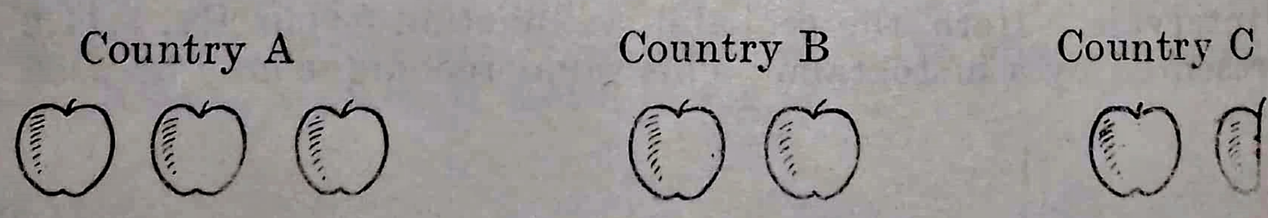


Fig. 3.6.

Exercises

3.1. Define two stochastic variables in each of the following experiments and write down their probability functions.

(*a*) A balanced coin is tossed 3 times.

(*b*) A balanced die is rolled twice.

(*c*) From a set of 20 girls and 30 boys 2 students are selected at random one by one without replacement.

3.2. Find the probability function of a stochastic variable x = the number of aces in a hand of bridge (one bridge hand contains 13 cards selected at random from a well shuffled set of 52 cards).

3.3. Suppose that a machine produces 2% defective items. If X denotes the number of defectives in a lot of 50 items randomly selected from a day's production find the probability function for X.

3.4. Assuming the probability that a new born baby in a particular hospital is a boy is 1/2, find the probability function for a stochastic variable X=the number of boys in sets of 100 births. Also find the probability that there are at least 45 boys among the newborn babies, in a set of 100 births.

3.2. DISTRIBUTION FUNCTION OR CUMULATIVE DISTRIBUTION

It can be defined as a function $F(x)$=the probability that the stochastic variable assumes values less than or equal to a specified value x, $=P\{X \leqslant x\}$. This gives the probability that a stochastic variable X falls below a specified point on the real line $R=(-\infty,\infty)$. For example $F(5)$=probability that X assumes a value less than or equal to 5 (that is, $x \leqslant 5$). Whenever there is no confusion we will use the notations, $F(x_0)=P(X \leqslant x_0)$ or $F(x_0)=P\{x \leqslant x_0\}$ where x_0 is a specified value.

Ex. 3.2.1. *Let x be a discrete s.v. with the probability function $f(x)$ given as follows :*

x	$f(x)$
0	*1/8*
1	*2/8*
2	*3/8*
3	*1/8*
4	*1/8*

and $f(x)=0$ elsewhere.

Solution.

From the table we can form the cumulative probabilities as follows :

$$F(0)=P\{x \leqslant 0\}=\tfrac{1}{8}$$
$$F(1)=P\{x \leqslant 1\}=\tfrac{1}{8}+\tfrac{2}{8}=\tfrac{3}{8}$$
$$F(2)=P\{x \leqslant 2\}=\tfrac{1}{8}+\tfrac{2}{8}+\tfrac{3}{8}=\tfrac{6}{8}$$
$$F(3)=P\{x \leqslant 3\}=\tfrac{1}{8}+\tfrac{2}{8}+\tfrac{3}{8}+\tfrac{1}{8}=\tfrac{7}{8}$$
$$F(4)=P\{x \leqslant 4\}=\tfrac{1}{8}+\tfrac{2}{8}+\tfrac{3}{8}+\tfrac{1}{8}+\tfrac{1}{8}=\tfrac{8}{8}=1.$$

The cumulative distribution or the distribution function may be given as

$$F(x)=\begin{cases} 0 \text{ for } x<0 \text{ or for } -\infty<x<0 \\ \tfrac{1}{8} \text{ for } x=0 \quad \text{or for } 0 \leqslant x<1 \\ \tfrac{3}{8} \text{ for } x=1 \quad \text{or for } 1 \leqslant x<2 \\ \tfrac{6}{8} \text{ for } x=2 \quad \text{or for } 2 \leqslant x<3 \\ \tfrac{7}{8} \text{ for } x=3 \quad \text{or for } 3 \leqslant x<4 \\ 1 \text{ for } x=4 \quad \text{or for } 4 \leqslant x<\infty \end{cases}$$

Comments. In this example the various ways of writing the distribution function $F(x)$ are given. It may be noticed that the distribution function varies from 0 to 1.

If the distribution function $F(x)$ of Ex. 3.2.1 is represented graphically, we get a step function as shown in Fig. 3.7. It may be further noticed from the discussions so far that distribution functions of discrete stochastic variables are always step functions.

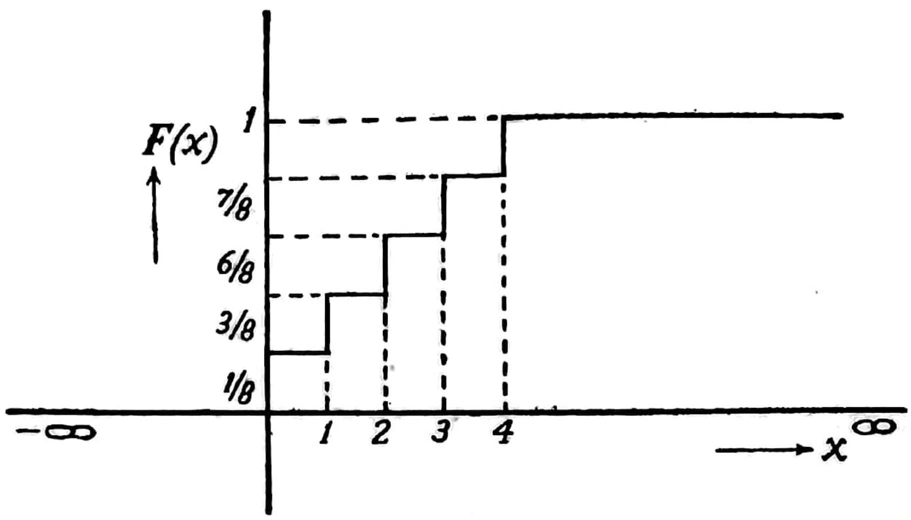


Fig. 3.7.

Evidently if the distribution function of a continuous Stochastic variable is represented by a graph, we can expect a smooth curve as shown in Fig. 3.8. The curve may be expected to

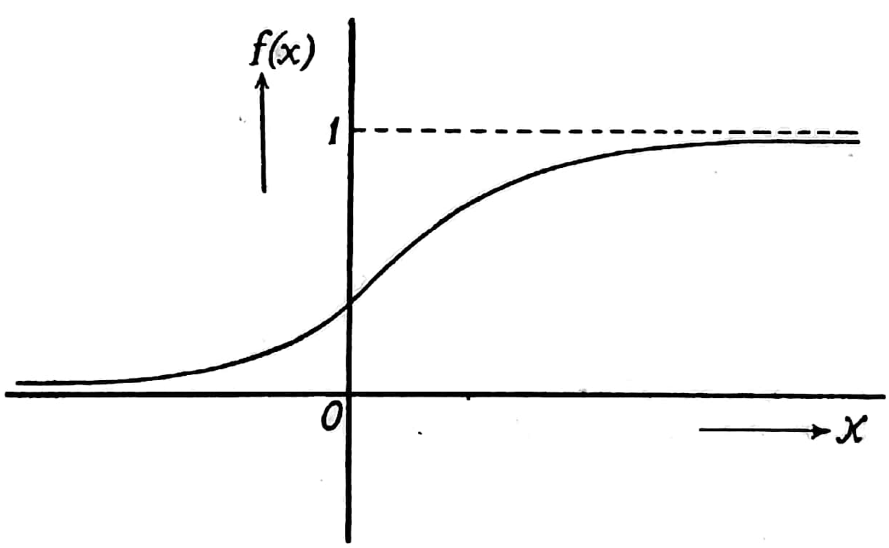


Fig. 3.8.

be smooth with the maximum ordinate equal to unity and with the minimum ordinate equal to zero. But the shape of the curve depends on the probability function of the *s.v.* The following properties are easily seen for a distribution function whether it is for a discrete or for a continuous *s.v.*

(1) $F(-\infty) = 0$

(2) $F(\infty) = 1$ (3·1)

(3) $F(a) \leqslant F(b)$ for $a < b$

In Ex. 3·2·1 for the probability that x lies between 1 and 4 (*i.e.* say, $1 < x \leqslant 4$) is given by $\frac{3}{8}+\frac{1}{8}+\frac{1}{8}=\frac{5}{8}$. This may be written

as $F(4)-F(1)=1-\frac{3}{8}=\frac{5}{8}$. In general the probability that x lies between two points x_0 and $x_0+\triangle x_0$ may be obtained by

$$F(x_0+\triangle x_0)-F(x_0).$$

(where x_0 and $x_0+\triangle x_0$ are two points. $x_0+\triangle x_0$ only means a point different from x_0 by a positive quantity $\triangle x_0$. For example 1 may be taken as x_0 and 4 may be taken as $x_0+\triangle x_0$ in Ex. 3·2·1). This may also be written as

$$\sum_{x_0<x\leqslant x_0+\triangle x_0} f(x) = P\{x_0<X\leqslant x_0+\triangle x_0\} \tag{3·2}$$

when X is a discrete stochastic variable.

This is equal to the area of the rectangles erected over the points from x_0 to $x_0+\triangle x_0$ in a histogram if the total area is assumed to be unity.

3.21. Density function. If $F(x)$ is a continuous function satisfying the conditions (3·1) and if $\frac{d}{dx}F(x)$ exists almost everywhere (except for a set of probability measure zero) and if $f(x)=\frac{d}{dx}F(x)$ then $f(x)$ may be defined as the density function for a continuous *s.v.* X for which the distribution is $F(x)$. It may be noticed that,

$$\int_{-\infty}^{x} f(x)\,dx=F(x)-F(-\infty)=F(x)-0=F(x). \tag{3·3}$$

Thus the probability function for a continuous *s.v.* X is often called the density function of X. The probability that $x_0<x\leqslant x_0+\triangle x_0$ may thus be obtained as,

$$\int_{x_0}^{x_0+\triangle x_0} f(x)\,dx=F(x_0+\triangle x_0)-F(x_0) \tag{3·4}$$

It may be noticed that when X is a continuous *s.v.* the probability that it takes a particular value, say x_1, is given by,

$$\int_{x_1}^{x_1} f(x)\,dx=0=P\{x=x_1\} \text{ when X is continuous.} \tag{3·5}$$

This shows that the probability measure $P\{x=x_1\}=0$ when X is continuous. In other words $f(x)$ can be a probability function for a continuous *s.v.*, X even if $f(x)$ has a discontinuity point provided the probability measure at that point is zero. It is intuitively evident that $f(x)$ can have a countable number of discontinuity points provided the total measure in those points is zero, that is, if $f(x)$ is continuous almost everywhere (a.e.) and if

$f(x) \geqslant 0$ for all x then $f(x)$ can be the probability function of a continuous *s.v.*, X, provided $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Further $\int_a^b f(x)\,dx$ is the area under the curve $f(x)$ between the ordinates at $x=b$ and $x=a$. For example if $f(x)$ is given as shown in Fig. 3·9 then $\int_a^b f(x)dx$ is the area of the shaded portion.

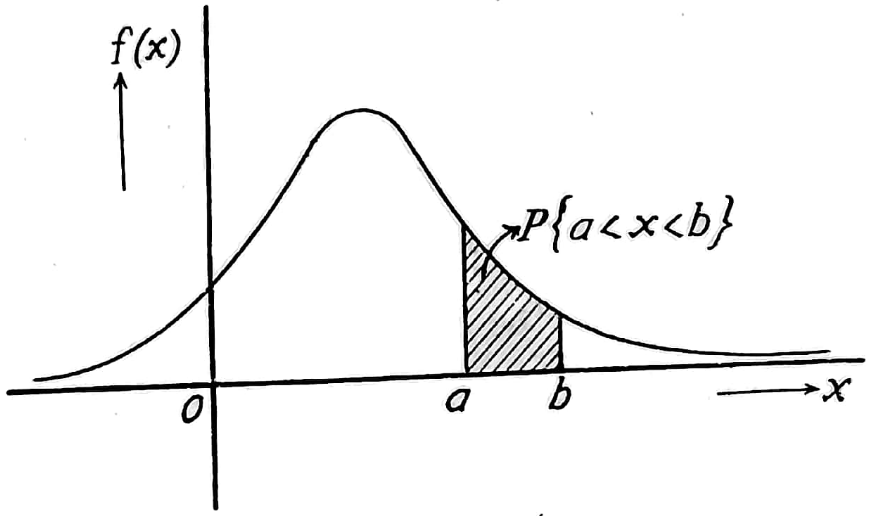


Fig. 3·9.

In general the probability that x is greater than or equal to x_0 is given

$$P\{x_0 \leqslant x \leqslant \infty\} = \sum_{x_0 \leqslant x \leqslant \infty} f(x) \text{ when } x \text{ is a discrete } s.v. \tag{3·6}$$

$$P\{x_0 \leqslant x \leqslant \infty\} = \int_{x_0}^{\infty} f(x)\,dx = P\{x_0 < x < \infty\}$$

when X is a continuous *s.v.*

The probability that $|x|$ is greater than x_0, where $|x|$ denotes the absolute value or magnitude of x, (without considering the sign), is given by

$$\sum_{-\infty \leqslant x < -x_0} f(x) + \sum_{x_0 < x \leqslant \infty} f(x) \text{ when X is a discrete } s.v.$$

and $\int_{-\infty}^{-x_0} f(x)\,dx + \int_{x_0}^{\infty} f(x)\,dx$ when X is a continuous *s.v.*

(3·7)

For a discrete *s.v.*, X, whose probability function $f(x)$ is represented by the histogram in Fig. 3·10, $P\{|x| \geqslant 1\}$ is given by the shaded area.

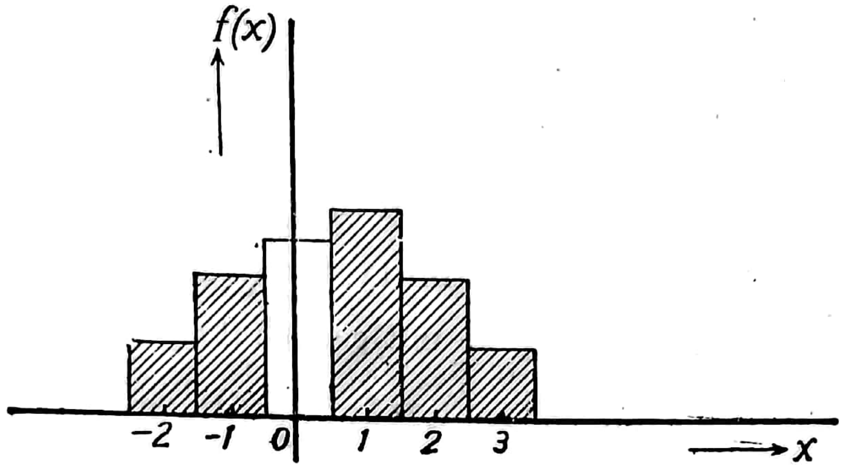


Fig. 3·10.

For a continuous *s.v.*, X whose probability function or density function is represented by the curve in Fig. 3·11, the $P\{|x| \geqslant 3\}$ is given by the shaded area.

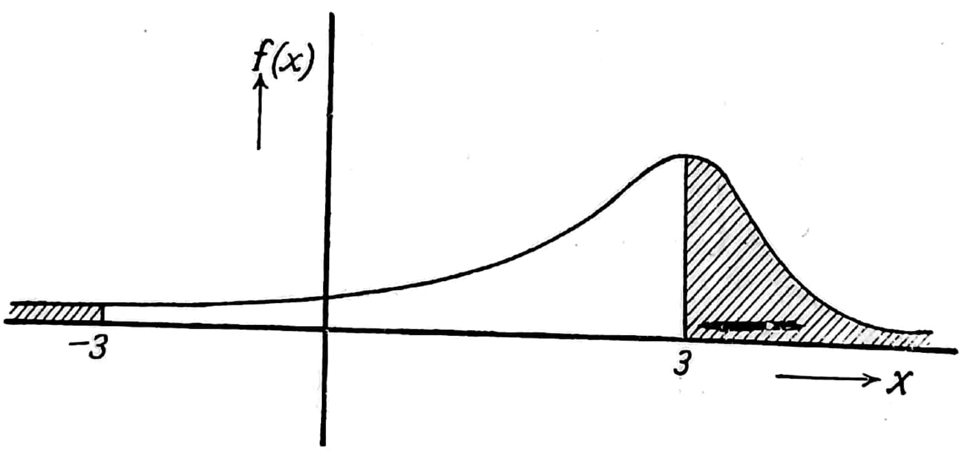


Fig. 3·11.

From the results discussed so far it is seen that if a function $f(x)$ is to be a probability function it should satisfy the following conditions :

1. $f(x) \geqslant 0$ for all x
2. $\sum\limits_{-\infty \leqslant x \leqslant \infty} f(x) = 1$ if X is a discrete *s.v.* (3·8)
3. $\int_{-\infty}^{\infty} f(x)\, dx = 1$ if X is a continuous *s.v.*

These conditions may be defined as the axioms or postulates for a probability function (*i.e.*, either the probability function for a discrete *s.v.* or the density function for a continuous *s.v.*) associated with a stochastic variable X.

Ex. 3.2.2. *$f(x)=\frac{1}{2x}$ for $x=1, 2, 3,$ and $f(x)=0$ elsewhere. Can this be a probability function for a s.v. X ?*

Solution. $f(1)=\frac{1}{2}$, $f(2)=\frac{1}{4}$, $f(3)=\frac{1}{6}$.

$\therefore$ $f(x)\geqslant 0$ for all x.

$$\sum_{-\infty\leqslant x\leqslant\infty} f(x)=\sum_{1\leqslant x\leqslant 3} f(x)=f(1)+f(2)+f(3)$$

$$=\frac{1}{2}+\frac{1}{4}+\frac{1}{6}=\frac{11}{12}\neq 1$$

Hence $f(x)$ is not a probability function.

Comments. It may be noticed that

$$f(x)=\frac{1}{2^x} \text{ for } x=1, 2, 3 \ldots\ldots$$

and is zero elsewhere, satisfies the conditions for a probability function, since $\sum_{-\infty\leqslant x\leqslant\infty} f(x)=\sum_{1\leqslant x\leqslant\infty}\frac{1}{2^x}=1$ and $f(x)\geqslant 0$ for all x. The *s.v.*, X may be easily seen to be discrete since x takes only individually distinct values.

Ex. 3.2.3. *A function $f(x)$ is given as follows*

$$f(x)=\begin{cases} x \text{ for } 0<x\leqslant 1 \\ \dfrac{3-x}{4} \text{ for } 1<x\leqslant 3 \\ =0 \text{ elsewhere.} \end{cases}$$

Can $f(x)$ be a probability function ? If so find the distribution function.

Solution. Evidently $f(x)\geqslant 0$ for all x.

$$\int_{-\infty}^{\infty} f(x)\,dx=\int_{-\infty}^{0} 0.dx+\int_{0}^{1} x.dx+\int_{1}^{3}\frac{3-x}{4}dx+\int_{3}^{\infty} 0\,dx$$

$$=0+\left[\frac{x^2}{2}\right]_0^1+\frac{1}{4}\left[3x-\frac{x^2}{2}\right]_1^3+0=1$$

$\therefore$ $f(x)$ is a density function.

By definition the distribution function

$$F(x)=\int_{-\infty}^{x} f(x)\,dx$$

$\therefore$ For any x such that $-\infty<x\leqslant 0$

$$F(x)=\int_{-\infty}^{x} 0.dx=0$$

For any x where $0<x\leqslant 1$

$$F(x)=\int_{-\infty}^{0} 0.dx+\int_{0}^{x} x.dx=\frac{x^2}{2}$$

For any x where $1<x\leqslant 3$

$$F(x)=\int_{-\infty}^{0} 0.dx+\int_{0}^{1} x.dx+\int_{1}^{x}\frac{3-x}{4}dx$$

$$=0+\frac{1}{2}+\frac{1}{4}\left[3x-\frac{x^2}{2}\right]_1^x$$

$$=-\frac{1}{8}+\frac{1}{4}\left(3x-\frac{x^2}{2}\right)$$

For any x where $3\leqslant x<\infty$

$$F(x)=\int_{-\infty}^{0} 0.dx+\int_{0}^{1} x\,dx+\int_{1}^{3}\frac{3-x}{4}dx+\int_{3}^{x} 0\,dx$$

$=1.$

$\therefore$ The distribution function $F(x)$ may be given as

$$F(x)=\begin{cases} 0 \text{ for } -\infty<x\leqslant 0 \\ \dfrac{x^2}{2} \text{ for } 0<x\leqslant 1 \\ -\dfrac{1}{8}+\dfrac{1}{4}\left(3x-\dfrac{x^2}{2}\right) \text{ for } 1<x\leqslant 3 \\ 1 \qquad \text{for } 3\leqslant x<\infty. \end{cases}$$

Comments. In this example the form of the density function is different in different intervals. Therefore in order to find out the distribution function we had to consider x in the various intervals, separately. If $f(x)$ had only one form throughout the interval $(-\infty, \infty)$ then $\int_{-\infty}^{x} f(x)\,dx$ can give $F(x)$ directly in the interval $(-\infty, \infty)$.

Ex. 3.2.4. *Given the density function*

$$f(x)=\begin{cases} k(2-x) \text{ for } 0<x<2 \\ 0 \quad \text{elsewhere} \end{cases}$$

find k.

Solution. If $f(x)$ is a density function

$$F(\infty)=\int_{-\infty}^{\infty} f(x)dx=1$$

i.e.,
$$\int_{-\infty}^{\infty} f(x)dx=0+\int_{0}^{2} k(2-x)dx=k\left[2x-\frac{x^2}{2}\right]_0^2$$

$$=2k$$

$\therefore \qquad 2k=1$

or $\qquad k=1/2.$

Comments. If instead of a density function a probability function $f(x)$ is given for a discrete *s.v.*X together with the range of X, then we can use the result that $\sum\limits_{-\infty<x<\infty} f(x)=1$ and any unknown quantity can be evaluated. These discussions indicate that to any function $f(x)$ which is non-negative where $\sum\limits_{-\infty<x<\infty} f(x)$ or $\int_{-\infty}^{\infty} f(x)dx$ exists, we can associate a probability function as follows.

Let $\sum\limits_{-\infty<x<\infty} f(x)$ or $\int_{-\infty}^{\infty} f(x)$ be equal to k. Consider the function $\phi(x)=f(x)/k$ then $\phi(x)$ satisfies all the conditions for a probability function.

Ex. 3.2.4. *Given that*

$$f(x,\theta)=\theta e^{-\theta x} \textit{ for } 0<x<\infty \textit{ and}$$

$f(x,\theta)=0$ elsewhere, where $\theta>0$ is a constant; can this be a probability function?

Solution. $\int_{-\infty}^{\infty} f(x,\,\theta)dx=0+\int_{0}^{\infty} \theta e^{-\theta x}dx=\theta\int_{0}^{\infty} e^{-\theta x}\,dx$ (since θ is a constant)

$$=\theta\left[-e^{-\theta x}/\theta\right]_0^\infty=1.$$

Evidently $f(x,\theta)\geqslant 0$ and hence $f(x,\theta)$ can be a density function irrespective of the value of θ provided $\theta>0$.

Comments. In this density function, θ is a constant with respect to the *s.v.*X. Such constants in probability functions are called parameters. In this example there is one parameter θ. In general a probability function may be denoted by $f(x,\theta)$ where θ denotes all the parameters in the given probability function. Hereafter this general notation will be used for probability functions whenever it is convenient. If the parameters in a probability function are given then the probability function is completely specified. For example if the above example was given as,

$$f(x)=5e^{-5x} \text{ for } 0<x<\infty$$

and $$f(x)=0 \text{ elsewhere,}$$

then there is no parameter or $f(x)$ is completely specified. For various values of θ we get a number of probability functions, all having the same functional form. In general, we can say that $f(x,\theta)$ denotes a family of probability functions.

Exercises

3·5. A consignment of 30 electric bulbs contain 10 defective ones. A random sample of 12 is selected from the set. If X denotes the number of defectives in the sample find (1) the probability function for X, (2) the distribution function for X.

3·6. Can the following functions be probability functions ? If so, find the corresponding distribution functions.

(*a*) $$f(x)=\begin{cases} 1/4 & \text{for } x=1 \\ 1/2 & \text{for } x=2 \\ 0 & \text{elsewhere.}\end{cases}$$

(*b*) $$f(x)=\begin{cases} 1/3 & \text{for } x=-1 \\ 1/3 & \text{for } x=0 \\ 1/3 & \text{for } x=5 \\ 0 & \text{elsewhere.}\end{cases}$$

(*c*) $$f(x)=\begin{cases} 3x & \text{for } 0<x<1 \\ 0 & \text{elsewhere.}\end{cases}$$

(*d*) $$f(x)=\begin{cases} x & \text{for } 0<x<1 \\ 2-x & \text{for } 1\leqslant x<2 \\ 0 & \text{elsewhere.}\end{cases}$$

3·7. In problem 3·6 represent the discrete distribution, if there is any, by bar diagrams and histograms and the continuous distribution, if there is any, by a curve. Also sketch the distribution functions in these cases.

3·8. Evaluate k, if the following functions are probability functions.

(a) $f(x)=\begin{cases} k/2 \text{ for } x=0 \\ k/4 \text{ for } x=2 \\ k/4 \text{ for } x=5 \\ 0 \text{ elsewhere.} \end{cases}$

(b) $f(x)=\begin{cases} k(1-x^2) \text{ for } 0<x<1 \\ 0 \text{ elsewhere.} \end{cases}$

(c) $f(x)=\begin{cases} 2k\theta e^{-\theta x} \text{ for } 0<x<\infty \text{ where } \theta>0 \text{ and is a constant} \\ 0 \text{ elsewhere.} \end{cases}$

3·9. If a stochastic variable X has a probability function $f(x)$ as given elow, evaluate and illustrate graphically

(a) $P\{ x \geqslant 3 \}$,

(b) $P\{ |x| < 1{\cdot}5\}$,

(c) $P\{ 1<x<3\}$.

$$f(x)=\begin{cases} \frac{x}{2} & \text{for } 0<x\leqslant 1 \\ \frac{3-x}{4} & \text{for } 1<x\leqslant 2 \\ \frac{1}{4} & \text{for } 2<x\leqslant 3 \\ \frac{4-x}{4} & \text{for } 3<x<4 \\ 0 & \text{elsewhere.} \end{cases}$$

3·10. If a stochastic variable X is given by

$$f(x)=\begin{cases} 1/4 \text{ for } x=-2 \\ 1/4 \text{ for } x=0 \\ 1/2 \text{ for } x=5 \\ 0 \quad \text{elsewhere.} \end{cases}$$

Evaluate and illustrate by a histogram

(a) $P\{x \leqslant 0\}$,

(b) $P\{x < 0\}$,

(c) $P\{ |x| \geqslant 2\}$,

(d) $P\{0 \leqslant x \leqslant 10\}$.

3·11. If the distribution function $F(x)$ is given to be

$$F(x)=\begin{cases} 2x^2/5 & \text{for } 0<x\leqslant 1 \\ -3/5+2(3x-x^2/2)/5 & \text{for } 1<x\leqslant 2 \\ 1 & \text{for } x>2 \end{cases}$$

Find the density function $f(x)$ and sketch $f(x)$ and $F(x)$.

3·12. If $f(x)$ is given as,

$$f(x)=\begin{cases} 1/8 \text{ for } x=-2 \\ 2/8 \text{ for } x=-1 \\ 3/8 \text{ for } x=0 \\ 2/8 \text{ for } x=2 \\ 0 \quad \text{elsewhere,} \end{cases}$$

evaluate (1) $F(x)$, (2) $F(0)-F(-1)$, (3) $P\{x \geqslant 0\}$.

3·13. Construct two examples, if possible, of a probability density function which has different functional forms in the following intervals, $-\infty < x \leqslant 1$, $1 < x \leqslant 2$, $2 < x \leqslant 3$, $3 < x < \infty$.

3·14. Construct two examples of a distribution function of a continuous *s.v.* X which has different functional forms in successive intervals. of the following intervals. $-\infty < x \leqslant 0$, $0 < x \leqslant 2$, $2 < x < \infty$.

3.3. MATHEMATICAL EXPECTATION

The mathematical expectation of a function $\psi(X)$ (where ψ is a Greek letter called psi) of a stochastic variable X is defined as

$$E[\psi(X)] = \sum_{-\infty < x < \infty} \psi(x) f(x) \text{ when X is discrete}$$

$$= \int_{-\infty}^{\infty} \psi(x) f(x)\, dx \text{ when X is continuous.} \qquad (3.9)$$

Here $\psi(X)$ is a function of X, like X^2, X, $X(X-1)$ etc. Any function of a *s.v*, X need not be a *s.v*. But in our discussion we will consider only functions which are *s.v*'s. Whenever there is no confusion $E[\psi(X)]$ will be written as $E\,\psi(X)$. $E[\psi(X)]$ is read as expectation of $\psi(X)$ where E denotes 'mathematical expectation'. E may also be considered to be an operator like the differential operator D, the integral operator $\int$ etc. This operator E plays a vital role in statistical analysis. When $\psi(X)=X$,

$$E(X) = \sum_{-\infty < x < \infty} x \; f(x) \quad \text{when X is discrete}$$

$$= \int_{-\infty}^{\infty} x\, f(x)\, dx \quad \text{when X is continuous.} \qquad (3.10)$$

E(X) is sometimes called the mean value of X. E(X) is a concept of average which may be seen from the following example.

Ex. 3.3.1. *A discrete s.v X takes the values $x_1, x_2, \ldots, x_n$ with probabilities $\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}$ and other values with probability zero. What is E(X) ?*

Solution. By definition

$$E(X) = x_1 \cdot \frac{1}{n} + x_2 \cdot \frac{1}{n} + \cdots + x_n \cdot \frac{1}{n}$$

$$= (x_1 + x_2 + \cdots + x_n)/n.$$

Ex. 3.3.2. *Consider the following situation. A person gets a sum of money equal to the square of the number that appears on the face when a balanced die with the faces marked 1, 2, 3,......6 is rolled. This die is rolled a number of times. In the long run, i.e., when the number of trials tends to infinity or approximately, when this game is repeated for a very very large number of times, how much money can he expect on the average per game ?*

Solution.

When a balanced die is rolled the numbers 1, 2, 3, 4, 5, 6 can come with probability 1/6, 1/6,...1/6. If the number 4 comes at a particular trial he gets \$ 4^2=\$ 16. When the game is repeated a large number of times the relative frequencies of 1, 2, 3,...6 approximate to 1/6, 1/6,...1/6. Therefore the total amount of money that he gets on the average is

$$=\$(1^2\times\tfrac{1}{6}+2^2\times\tfrac{1}{6}+\cdots+6^2\times\tfrac{1}{6})$$
$$=\$\tfrac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2)=\$15.17.$$

Comments. It is easily seen that X is a *s.v.*, which is defined as the square of the number rolled then the amount the person gets is $E(X)=\Sigma\, x\, f(x)=1^2\times\frac{1}{6}+2^2\times\frac{1}{6}+\cdots+6^2\times\frac{1}{6}=15.17.$

From this example it is seen that E(X) is a concept of average.

Ex. 3.3.3. *Find E(X) where X is defined as follows*

$$f(x)=\begin{cases} 0 & \text{for } x<0 \\ x & \text{for } 0\leqslant x<1 \\ \dfrac{3-x}{4} & \text{for } 1\leqslant x<3 \\ 0 & \text{for } x\leqslant 3. \end{cases}$$

Solution.

$$E(X)=\int_{-\infty}^{\infty} x.f(x)dx=\int_{-\infty}^{0} x.0\,dx+\int_{0}^{1} x.x\,dx+\int_{1}^{3} x\cdot\frac{3-x}{4}dx$$
$$+\int_{3}^{\infty} x.0\,dx$$

$$=0+\left[\frac{x^3}{3}\right]_0^1+\frac{1}{4}\left[3\,\frac{x^2}{2}-\frac{x^3}{3}\right]_1^3+0=7/6.$$

3.3.1. Moments. E(X) is usually called the first moment about the origin and is usually denoted by μ_1' or μ (where μ is a Greek letter called mu and μ_1^1 is read as mu one prime).

i.e., $$E(X)=\mu_1^1=\mu.$$

In general the r^{th} moment about the origin (sometimes called the r^{th} raw moment) is defined as $E(X^r)$ and is usually denoted by μ_r' (called mu r prime).

$$\mu_r' = E(X^r) = \sum_{-\infty < x < \infty} x^r \ f(x) \quad \text{when X is discrete} \tag{3·11}$$

$$= \int_{-\infty}^{\infty} x^r f(x)\, dx) \text{ when X is continuous}$$

i.e., First moment about the origin

$= \mu_1' = E(X)$.

Second moment about the origin $= \mu'_2 = E(X^2)$ etc.

Theorem 3.3.1. $E(c) = c$ where c is a constant with respect to a *s.v.*X.

Proof. By definition $E(c)$

$$= \int_{-\infty}^{\infty} c \,.\, f(x)dx \qquad \text{when X is continuous}$$

$$= c \,.\int_{-\infty}^{\infty} f(x)dx$$

$$= c \,.\, 1 \qquad \left(\text{since} \int_{-\infty}^{\infty} f(x)dx = 1\right)$$

$$= c.$$

The proof when X is discrete is left to the reader.

Theorem 3.3.2. $E[c \,.\, \psi(X)] = c \,.\, E[\psi(X)]$ where c is a constant.

Proof. $E[c \,.\, \psi(X)]$

$$= \sum_{-\infty < x < \infty} c \,.\, \psi(x) f(x) \qquad \text{when X is discrete}$$

$$= c \sum_{-\infty < x < \infty} \psi(x) f(x) \qquad (\text{since } c \text{ is a constant})$$

$$= c\, E[\psi(X)] \qquad (\text{by definition}) \tag{3·12}$$

The proof when X is continuous is left to the reader.

Corollary. $E[cX] = c \,.\, E(X) = c \,.\, \mu'_1$ where c is a constant. (3·13)

Theorem 3.3.3. $E(aX+b) = a\, E(X) + b$, where a and b are constants. (3·14)

Proof. $E(aX+b)$

$$=\int_{-\infty}^{\infty} (ax+b)\, f(x)dx \quad \text{when X is continuous}$$

$$=\int_{-\infty}^{\infty} ax\,.\, f(x)dx+\int_{-\infty}^{\infty} bf(x)dx$$

$$=a\,.\int_{-\infty}^{\infty} x\,.\, f(x)dx+b\int_{-\infty}^{\infty} f(x)dx$$

(since a and b are constants)

$$=a\,E(X)+b,$$

The proof when X is discrete is left to the reader.

Corollary 1. $E[a\,\psi(X)+b\,\phi(X)]=a\,E\psi(X)+bE\phi(X)$ where a and b are constants, $\psi(X)$ and $\phi(X)$ are functions of X.

Corollary 2. $E[X-E(X)]=E[X-\mu'_1]=0.$ (3·15)

3·32. Central Moments. The r^{th} moment about a point c may be defined as $E(X-c)^r$ where c is a constant. When c is $E(X)=\mu$ then the r^{th} moment about E(X) is obtained. This is usually called the r^{th} central moment and is usually denoted by μ_r.

$$\mu_r=E(X-\mu)^r$$

where $\mu=E(X)$

$$=\int_{-\infty}^{\infty} (x-\mu)^r f(x)dx \quad \text{when X is continuous}$$

$$=\sum_{-\infty<x<\infty} (x-\mu)^r f(x) \quad \text{when X is discrete} \qquad (3·16)$$

For example

$$\mu_2=E(X-\mu)^2$$
$$=E[X^2-2\mu X+\mu^2]$$
$$=E(X^2)-2\mu E(X)+E(\mu^2)$$
$$=E(X^2)-\mu^2. \qquad (\text{since } E(X)=\mu \text{ and } E\mu^2=\mu^2)$$

(3·17)

μ_2 is sometimes called the variance of the stochastic variable X [Var (X)]. The positive square root of the variance (*i.e.*, $\sqrt{\mu_2}$) is called the standard deviation of the stochastic variable X and is usually denoted by σ.

$$\sigma=\sqrt{\mu_2}=[E(X-\mu)^2]^{1/2}$$

$$=\{E[X-E(X)]^2\}^{1/2}=[\text{Var}(X)]^{1/2}.$$

It may be noticed that $[E(X-\mu)^2]^{1/2}$ need not be equal to $E(X-\mu)$. The standard deviation may also be called the root mean square deviation when the square deviations are taken from $E(X)=\mu$.

Theorem 3.3.4. $\text{Var}(aX+b)=a^2 \cdot \text{Var}(X)$, where a and b are constants.

Proof. By definition

$$\begin{aligned}\text{Var}(aX+b)&=E[aX+b-aE(X)-b]^2\\&=E[a\{X-E(X)\}]^2\\&=E\, a^2[X-E(X)]^2\\&=a^2\, E[X-E(X)]^2 \qquad \text{(since } a \text{ is a constant]}\\&=a^2 \cdot \text{Var}(X)\end{aligned} \tag{3·18}$$

Corollary. $\text{Var}(X+b)=\text{Var}(X)$ where b is a constant.

Theorem 3.3.5. $\mu_r=\mu'_r-\binom{r}{1}\mu'_{r-1}\cdot\mu+\binom{r}{2}\mu'_{r-2}\cdot\mu^2-\ldots\ldots+(-1)^r\mu^r$ where μ_r is the r^{th} central moment and μ'_r is the r^{th} raw moment and μ is the first raw moment or $E(X)=\mu=\mu'_1$.

Proof. By definition

$$\mu_r=E[X-\mu]^r$$

where $\mu=E(X)$ and r is a positive integer.

But $(X-\mu)^r=X^r-\binom{r}{1}X^{r-1}\cdot\mu+\binom{r}{1}X^{r-2}\mu^2-\ldots\ldots+(-1)^r\mu^r.$

$$\therefore \quad \mu_r=E(X-\mu)^r=E(X^r)-\binom{r}{1}\mu\cdot E(X^{r-1})$$

$$+\binom{r}{2}\mu^2\cdot E(X^{r-2})-\ldots\ldots+(-1)^r\mu^r E(1)$$

$$=\mu'_r-\binom{r}{1}\mu'_{r-1}\cdot\mu+\binom{r}{2}\mu'_{r-2}\cdot\mu^2-\ldots\ldots+(-1)^r\mu^r \tag{3·19}$$

Corollary 1. $\mu_2=\mu'_2-\mu^2.$

Corollary 2. $\mu_3=\mu'_3-3\mu'_2\cdot\mu+3\mu'_1\cdot\mu^2-\mu^3$

$$=\mu'_3-3\mu'_2\cdot\mu+2\mu^3 \tag{3·20}$$

There are other types of moments which will be briefly discussed in the following sections.

3.33. Absolute Moments. The r^{th} absolute moment M_r about a constant c is defined as

$$M_r = E \mid X - c \mid^r$$

where $\mid X-c \mid$ denotes the absolute value or the magnitude of $X-c$. By this definition the first, second, third etc., absolute moments about c, are

$$M_1 = E \mid X - c \mid \tag{3·21}$$

$$M_2 = E \mid X - c \mid^2 \tag{3·22}$$

$$M_3 = E \mid X - c \mid^3 \text{ etc.} \tag{3·23}$$

when c is $E(X)=\mu$ the absolute moments are said to be absolute moments about μ. For example the second absolute moment about μ is $M_2 = E \mid X-\mu \mid^2 = E(X-\mu)^2$

$$= \mu_2 = \text{Var}(X)$$

(It may be noticed that $(X-\mu)^2 = (\mu - X)^2 = \mid X - \mu \mid^2$ since we are dealing only with real quantities.)

The first absolute moment from c is called the mean absolute deviation or mean deviation of the *s.v.*, X from c.

Mean deviation (abbreviation, M.D.) $= M_1 = E \mid X - c \mid$

It is evident that M_1 need not be equal to $\sqrt{M_2}$.

Ex. 3.33.1. *Find (1) the mean deviation from $\mu = E(X)$, (2) the standard deviation, (3) the third raw moment, (4) the third absolute moment from μ, for the following probability function.*

$$f(x) = \begin{cases} 1/4 & \text{for } x=0 \\ 1/2 & \text{for } x=1 \\ 1/4 & \text{for } x=2 \\ 0 & \text{elsewhere} \end{cases}$$

Solution. $\mu = E(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1.$

(1) The mean deviation from $\mu = M_1 = E \mid X - \mu \mid$

$$= \mid 0-1 \mid \times \tfrac{1}{4} + \mid 1-1 \mid \times \tfrac{1}{2} + \mid 2-1 \mid \times \tfrac{1}{4}$$

$$= \tfrac{1}{4} + 0 = \tfrac{1}{4} = 1/2.$$

(2) The standard deviation $= \sqrt{\mu_2}$

$$= [E(X-\mu)^2]^{\frac{1}{2}}$$

$$\mu_2 = (0-1)^2 \times \tfrac{1}{4} + (1-1)^2 \times \tfrac{1}{2} + (2-1)^2 \times \tfrac{1}{4}$$

$$= \tfrac{1}{4} + \tfrac{1}{4} = 1/2$$

$\therefore$ The standard deviation

$$\sigma = \sqrt{1/2} = 1/\sqrt{2}.$$

(3) The third raw moment

$$= \mu_3' = E(X^3)$$

$$= 0^3 \times \tfrac{1}{4} + 1^3 \times \tfrac{1}{2} + 2^3 \times \tfrac{1}{4} = 5/2$$

(4) The third absolute moment about $\mu = M_3 = E\mid X-\mu\mid^3$

$$=\mid 0-1\mid^3\times\tfrac{1}{4}+\mid 1-1\mid^3\times\tfrac{1}{2}+\mid 2-1\mid^3\times\tfrac{1}{4}$$

$$=1^3\times\tfrac{1}{4}+0\times\tfrac{1}{2}+1^3\times\tfrac{1}{4}=1/2.$$

Ex. 3·33·2. *Find (1) the mean deviation from θ, (2) standard deviation and (3) the second raw moment for the following probability function*

$$f(x)=\frac{1}{\theta} \text{ for } 0<x<\theta$$

$$=0 \text{ elsewhere, where } \theta>0 \text{ and is a parameter.}$$

Solution. (1) By definition the mean deviation from θ is

$$E\mid x-\theta=\int_0^\theta \mid x-\theta\mid \frac{1}{\theta}\,dx$$

$$=\frac{1}{\theta}\int_0^\theta \mid x-\theta\mid dx$$

But $x<\theta$ always ; therefore $\mid x-\theta\mid=\theta-x$.

$$\therefore \quad E\mid x-\theta\mid=\frac{1}{\theta}\int_0^\theta(\theta-x)dx=\frac{1}{\theta}\left[\theta x-\frac{x^2}{2}\right]_0^\theta$$

$$=\frac{1}{\theta}\left[\theta^2-\frac{\theta^2}{2}\right]=\frac{1}{\theta}\theta^2/2=\theta/2.$$

(2) Standard deviation

$$=\sqrt{\mu_2}=\{E[X-E(X)]^2\}^{\frac{1}{2}}$$

$$E(X)=\int_0^\theta x\frac{1}{\theta}\,dx=\frac{1}{\theta}\int_0^\theta x\,dx=\theta/2.$$

$$\therefore \quad \mu_2=E(X-\theta/2)^2=EX^2-(\theta/2)^2$$

and

$$E(X^2)=\int_0^\theta x^2\frac{1}{\theta}\,dx=\left[\frac{1}{\theta}\,\frac{x^3}{3}\right]_0^\theta=\theta^2/3$$

$$\therefore \quad \mu_2=\frac{\theta^2}{3}-\frac{\theta^2}{4}=\theta^2/12.$$

$\therefore$ The standard deviation

$$\sigma=\sqrt{\mu_2}=\sqrt{\theta^2/12}=\theta/\sqrt{12}$$

(3) The second raw moment

$$=E(X^2)$$
$$=6^2/3$$

Theorem 3.3.6. $$E\left(\frac{X-\mu}{\sigma}\right)=0$$

and $$\text{Var}\left(\frac{X-\mu}{\sigma}\right)=1$$

where $$\mu=E(X)$$

and $$\sigma=\sqrt{\mu_2}$$

Proof. Let $$Y=\frac{X-\mu}{\sigma}$$

$$E(Y)=E\left(\frac{X-\mu}{\sigma}\right)=$$

$$=\frac{1}{\sigma}\ E(X-\mu)=\frac{1}{\sigma}[E(X)-E(\mu)]$$

$$=\frac{1}{\sigma}[\mu-\mu]=0 \tag{3·24}$$

$$\text{Var}\ (Y)=E[Y-E(Y)]^2$$

$$=E\left[\frac{X-\mu}{\sigma}-0\right]^2$$

$$=\frac{1}{\sigma^2}\ E(X-\mu)^2=\frac{1}{\sigma^2}\ [\ \text{Var}\ (X)]$$

$$=\frac{\sigma^2}{\sigma^2}=1. \tag{3·25}$$

A stochastic variable whose expected value is zero and whose variance is unity is called a standardized stochastic variable.

3·34. Factorial moments. The r^{th} factorial moment of a stochastic variable X is defined as the expected value of

$X(X-1)\ldots(X-r+1)$ and is usually denoted by $\mu_{[r]}$.

$$\mu_{[r]}=E[X(X-1)(X-2)\ldots(X-r+1)]$$

$$=\int_{-\infty}^{\infty} x(x-1)\ldots(x-r+1)f\ (x)dx$$

when X is continuous (3·26)

$$=\sum_{-\infty<x<\infty} x(x-1)\ldots(x-r+1)\ f\ (x)$$

when X is discrete.

Ex. 3·34·1. *Find the second factorial moment of the stochastic variable X whose probability function is defined as*

$$f(x)=\begin{cases} \dfrac{1}{\theta} \text{ for } 0<x<\theta, & \theta>0 \text{ is parameter} \\ 0 \text{ elsewhere.} \end{cases}$$

This is known as a uniform or rectangular distribution with one parameter.

Sol. By definition the second factorial moment

$$\mu_{[2]}=E[X(X-1)]$$

$$=\int_{-\infty}^{\infty} x(x-1)f(x)dx$$

$$=\int_{0}^{\theta} x(x-1)\frac{1}{\theta}\, dx$$

$$=\frac{1}{\theta}\int_{0}^{\theta} (x^2-x)dx$$

$$=\frac{1}{\theta}\left[\frac{\theta^3}{3}-\frac{\theta^2}{2}\right]=\frac{\theta^2}{3}-\frac{\theta}{2}$$

Comments. $\mu_{[2]}$ may be expressed in terms of μ'_2 and μ'_1

$$\mu_{[2]}=EX(X-1)=E(X^2-X)=EX^2-EX$$

$$=\mu'_2-\mu'_1$$

$$\mu'_2=\mu_2+(\mu'_1)^2$$

$$\therefore \quad \mu_2=\mu_{[2]}+\mu_1'-(\mu_1')^2$$

$$=\mu_{[2]}+\mu-\mu^2$$

where
$$\mu=\mu_1'=E(X) \tag{3·27}$$

Ex. 3.34.2. *Find the first and second factorial moments for the stochastic variable whose probability function is given as*

$$f(x)=\begin{cases} \dfrac{\lambda^x}{x!}\, e^{-\lambda} \text{ for } x=0, 1, 2, \ldots\infty, \ \lambda>0 \text{ is a parameter.} \\ 0 \text{ elsewhere} \end{cases}$$

This distribution is known as Poisson distribution with the parameter. λ (lamda).

Sol. The first factorial moment is

$$\mu_{[1]}=E(X)=\mu_1'=\mu$$

$$\therefore \quad \mu_{[1]}=\sum_{-\infty<x<\infty} x\, f(x) = \sum_{x=0}^{\infty} x\, \frac{\lambda^x}{x\,!}\, e^{-\lambda}$$

$$=e^{-\lambda}\sum_{x=1}^{\infty} x\, \frac{\lambda^x}{x\,!}$$

(when $x=0$ the corresponding term is zero)

$$= e^{-\lambda} \cdot \lambda \cdot \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)\,!}$$

$$= e^{-\lambda} \cdot \lambda \left[1 + \frac{\lambda}{1\,!} + \frac{\lambda^2}{2\,!} + \frac{\lambda^3}{3!} + \dots \right] \tag{3·28}$$

$$= e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \lambda.$$

$$\left(\text{since } e^{\lambda} = 1 + \frac{\lambda}{1\,!} + \frac{\lambda^2}{2\,!} + \dots \right)$$

The second factorial moment

$$\mu_{[2]} = \mathrm{E}X(X-1)$$

$$= \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x}{x\,!} e^{-\lambda}$$

$$= \sum_{x=2}^{\infty} x(x-1) \frac{\lambda^x}{x\,!} e^{-\lambda}$$

(The terms corresponding to $x=0$ and 1 are zeros).

$$= \lambda^2 e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)\,!}$$

$$= \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2 \tag{3·29}$$

Comments. It can be easily verified that, for discrete stochastic variables, factorial moments are usually easier to evaluate.

A stochastic variable X together with its probability function $f(x)$ is sometimes called the probability distribution of the stochastic variable X. This is evidently different from the distribution function which is the cumulative probability function. For example Ex. 3.34.1 may be stated as follows. Find the second factorial moment for the following distribution

$$X : f(x) = \begin{cases} \dfrac{1}{\theta} \text{ for } 0 < x < \theta, \quad \theta > 0 \\ 0 \text{ elsewhere} \end{cases}$$

This means that the stochastic variable X has the probability function $f(x)$ which is defined as above. In this distribution (stochastic variable togther with its probability function) there is one parameter θ.

We defined various types of moments. But these moments need not always exist. A moment is said to exist if it is a finite quantity. Consider the following example.

Ex. 3.34.3. *Evaluate $E(2^x)$ for the probability distribution*

$$X : f(x)=\begin{cases} \dfrac{1}{2^x} & \text{for } x=1, 2, 3,\ldots\ldots \\ 0 & \text{elsewhere} \end{cases}$$

Sol. By definition

$$E(2^x)=\sum_{-\infty<x<\infty} 2^x f(x)$$

$$=\sum_{1\leqslant x\leqslant\infty} 2^x \frac{1}{2^x}.$$

$$=2\times\frac{1}{2}+2^2\times\frac{1}{2^2}+\ldots\ldots$$

$$=1+1+1+\ldots=\infty \tag{3·30}$$

Evidently the series $1+1+\ldots\ldots$ does not converge and therefore $E(2^x)$ for this distribution does not exist.

Comments. A series $a_1+a_2+\ldots$ is said to be convergent if

$$a_1+a_2+\ldots+a_n\to k \text{ as } n\to\infty$$

where k is a finite quantity. Otherwise the series is not convergent. If

$$a_1+a_2+\ldots+a_n\to\pm\infty \text{ as } n\to\infty$$

then the series is said to be divergent. Therefore depending upon the probability, distribution, moments may or may not exist.

3.35. Moment Generating Functions. If X is a stochastic variable then $E(e^{tX})$ is called the moment generating function (abbreviation M.G.F.) of X, where t is an arbitrary real constant with respect to the stochastic variable X. It will be seen in the following discussion that this M.G.F. gives the various raw moments and hence it is called a M.G.F. It is usually denoted by $M_X(t)$.

$$\therefore \quad M_X(t)=E(e^{tX})=E\left[1+tX+t^2\frac{X^2}{2!}+\frac{t^3}{3!}X^3+\ldots\right]$$

$$=E(1)+t\,E(X)+\frac{t^2}{2!}E(X^2)+\ldots$$

$$=1+t\,\mu_1'+\frac{t^2}{2!}\mu_2'+\frac{t^3}{3!}\mu_3'+\ldots \tag{3·31}$$

i.e., the coefficient of $\dfrac{t^r}{r!}$ in $M_X(t)$ gives the r^{th} raw moment μ_r' or the various raw moments $\mu_1', \mu_2', \ldots\ldots$ are obtained as the coefficients of

$$\frac{t}{1!}, \frac{t^2}{2!}, \frac{t^3}{3!}, \ldots \text{if } M_X(t) \text{ exists.}$$

This M.G.F. $M_X(t)$ exists if the series

$$1+t\mu_1'+\frac{t^2}{2!}\mu_2'+\ldots\ldots$$

is convergent for some t.

Ex. 3.35.1. *Obtain the M.G.F. for the Binomial probability distribution with the parameters N and p. (The Binomial probability distribution with the parameters N and p is given as*

$$f(x,\theta)=\begin{cases}\binom{N}{x}p^x(1-p)^{N-x} & \text{for } x=1,2,\ldots N\\ 0 \text{ elsewhere} & 0<p<1\end{cases}$$

Here θ represents the two parameters N and p.

Sol. $M_X(t)$ for the Binomial variate X is

$$M_X(t)=E(e^{tX})=\sum_{x=0}^{N} e^{tx}\cdot\binom{N}{x}p^x(1-p)^{N-x}$$

$$=\sum_{x=0}^{N}\binom{N}{x}(pe^t)^x(1-p)^{N-x}$$

Let $1-p=q$ and $pe^t=p'$ then

$$M_X(t)=\sum_{x=0}^{N}\binom{N}{x}(p')^x q^{N-x}$$

(Since $(q+p')^N$ when expanded by the Binomial expansion gives the sum above)

$$=(q+p')^N$$

$$=(q+pe^t)^N. \tag{3·32}$$

Theorem 3.3.7. $M_{X+a}(t)=e^{ta}M_X(t)$ where a is a constant or the M.G.F. of a stochastic variable $Y=X+a$ is e^{ta} times the M.G.F. of the stochastic variable X.

Proof. $M_{X+a}(t)=E\,e^{t(X+a)}=E\,e^{tX+ta}$

$$=E\,e^{tX}e^{ta}=e^{ta}E\,e^{tX}$$

(since e^{ta} is a constant)

$$=e^{ta}M_X(t) \tag{3·33}$$

Corollary. $M_{X-\mu}(t)=e^{-t\mu}M_X(t)$ where $\mu=E(X)=\mu_1'$ (3·34)

This gives a relationship between the central moments and raw moments.

Theorem 3.3.8. $\mu_r'=\frac{d^r}{dt^r}M_X(t)|t=0$ (3·35)

or the r^{th} raw moment is obtained by differentiating $M_X(t)$, r times with respect to t and substituting $t=0$, or μ'_r is the r^{th} derivative at $t=0$ of $M_X(t)$.

The proof is left to the reader.

Hint. Consider the series $M_X(t)$ and differentiate.

Ex. 3.35.2. *Obtain the first moment $\mu=E(X)$ from the M.G.F. in Ex. 3.35.1.*

Sol. The M.G.F. in Ex. 3.35.1 is seen to be

$$M_X(t)=(q+pe^t)^N$$

where $$q=1-p$$

$$\therefore \quad \frac{d}{dt}M_X(t)=N(q+pe^t)^{N-1}pe^t$$

$$\therefore \quad \frac{d}{dt}M_X(t) \mid t=0$$

$$=N(q+p)^{N-1}p \quad \text{(since } e^t=1 \text{ when } t=0)$$

$$=Np \quad \text{(since } q+p=1)$$

$$\therefore \quad \mu_1'=\mu=Np. \tag{3·36}$$

Comments. Similarly other raw moments may be easily obtained by successive differentiation of $(q+pe^t)^N$ with respect to t.

Theorem 3.3.9.

$$M_{aX+b}(t)=e^{tb}\,M_X(ta) \tag{3·37}$$

where a and b are constants.

Proof. By definition,

$$M_{aX+b}(t)=E\,e^{t(aX+b)}=E\,e^{taX}\,e^{tb}$$

$$=e^{tb}\,E\,e^{ta\,.\,X}$$

$$=e^{tb}\,M_X\,(ta).$$

Corollary. $$M_{\frac{X-\mu}{\sigma}}(t)=e^{-\frac{t\mu}{\sigma}}\,M_X(t/\sigma) \tag{3·38}$$

This gives a relation between the raw moments of standardized variate and that of the original variate.

Sometimes the M.G.F. $M_X(t)$ of a *s.v.* X may not exist, but in such cases another function called the characteristic function $\phi_X(t)$ of a *s.v.* X exists. The characteristic function $\phi_X(t)$ of a *s.v.* X is defined as

$$\phi_X(t)=E\,e^{itX} \text{ where } i=\sqrt{-1}$$

and t is an arbitrary real constant. Here also it may be seen that the coefficient of $\frac{(it)^r}{r!}$ in the expansion of $\phi_X(t)$ gives the

rth raw moment μ_r'. So $\phi_X(t)$ also generates the moments. The characteristic function does exist always and it also generates the raw or crude moments.

So far we have been defining various types of moments, M.G.F. and characteristic function of a *s.v.* X. For a given probability function, $M_X(t)$ and $\phi_X(t)$ may be evaluated, if $M_X(t)$ exists. It is natural to ask the question whether there exist, more than one probability function corresponding to, a M.G.F. or a characteristic function ? The answer is no ! The M.G.F. and characteristic function uniquely determine a probability function. For a proof of this result the reader may refer to Mathematical Methods of Statistics by H. Cramer. The uniqueness theorem gives the uniqueness of the probability distribution, if the M.G.F. or the characteristic function is given.

Ex. 3.35.3. *The M.G.F. of a probability distribution is given to be* $\frac{1}{81}(2+e^t)^4$. *What is the probability distribution ?*

Sol. $$\frac{1}{81}(2+e^t)^4=\left(\frac{2}{3}+\frac{1}{3}e^t\right)^4$$

This is of the form $(q+pe^t)^N$ where $q=2/3$, $p=1/3$ and $N=4$. Since the *M.G.F.* uniquely determines the corresponding probability function, the corresponding distribution is a Binomial distribution with the parameters $N=4$ and $p=1/3$. (See Ex. 3.35.1)

Exercises

3·15. A balanced coin is thrown 100 times under similar experimental conditions. What is the expected number of heads ?

3.16. A balanced die is rolled. If a person receives $10 when the number 1 or 3 or 5 occurs and loses $8 when 2, or 4 or 6 occurs. How much money can he expect on the average per roll in the long run ?

3·17. The probabilities that a man fishing at a particular place will catch 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 fishes are 0.60, 0.20, 0 06, 0.04, 0.03, 0.02, 0.02, 0.01, 0.01, 0.01 respectively. What is the expected number of fish caught ?

3.18. Suppose that the probabilities that sets of 1, 2, 3, 4, 5 persons come to visit a particular art gallery are 0.20, 0.50, 0.20, 0.07, 0.03 respectively. What is the expected number of persons per set ?

3.19. Prove that (1) $E(X-c)^2$ is a minimum when $c=E(X)=\mu$

(2) $E\mid X-c\mid$ is a minimum when $c=M=$median)

(See section 3.43)

3.20. Find the expected value and standard deviation for the following distributions.

(1) $$f(x)=\begin{cases} 1/8 \text{ for } x=-2 \\ 2/8 \text{ for } x=-1 \\ 3/8 \text{ for } x=0 \\ 2/8 \text{ for } x=2 \\ 0 \text{ elsewhere} \end{cases}$$

(2) $f(x)=\theta e^{-\theta x}$ for $0< x < \infty$ where $\theta > 0$ is a constant

$=0$ elsewhere.

3.21. If cumulants or semi-invariants k_1, k_2, ... of a probability distribution are defined as $\log M_x(t)=1+k_1t+k_2t^2/2!+...$ or $\log \phi_x(t)=1+k_1(it)+k_2(it)^2/2!+...$ whenever $M_x(t)$ does not exist, where log denotes the natural logarithms, show that the first two cumulants are such that $k_1=\mu=E(X)$ and $k_2=\mu_2=\text{Var}(X)$. $k_1, k_2, ...$ are called semi-invariants because except for k_1 all other k's are invariant (does not vary) under a translation of the variable. In other words, $k_2, k_3, ...$ for the *s.v*, X are the same as those for the *s.v*, X+c where c is a constant.)

3.22. The factorial moment generating function of a probability distribution is defined as $E(t^X)$ where t is an arbitrary real constant.

That is,

$$F_x(t)=E(t^X)$$

Show that the r^{th} derivative of $F_x(t)$ with respect to t at $t=1$ gives the r^{th} factorial moment.

3.23. A Cauchy distribution is defined as,

$$f(x)=1/\pi(1+x^2) \text{ for } -\infty < x < \infty$$

Show that E(X) for this distribution does not exist. Does $M_x(t)$ exist for this distribution ? [See also bibliography (6)].

3.24. The following are some moment generating functions. Find the corresponding probability distribution by using the uniqueness result.

(*a*) $M_x(t)=(1/2+e^t/2)^{10}$;

(*b*) $M_x(t)=(1+2e^t)^4/3^4$;

(*c*) $M_x(t)=(2+3e^t)^3/125$.

3.25. Obtain the moment generating function of the following distribution.

$$(x)=\begin{cases} x & \text{for } 0 < x < 1 \\ 2-x & \text{for } 1 \leqslant x < 2 \\ 0 & \text{elsewhere.} \end{cases}$$

3.4. SOME USES OF MOMENTS

Moments are usually used to specify a probability distribution (this may be noticed from the uniqueness theorem), to locate a specified point, to measure the scatter or dispersion, to measure symmetry or skewness (lack of symmetry) and to measure Kurtosis or peakedness in a probability distribution. Some of these uses are discussed in the following notes.

3.41. Points of location. If a statistical population is specified by a stochastic variable X or by the probability distribution $f(x, \theta)$, we may be interested in a point, say c, such that $P\{x \leqslant c\}=p\%$ of the total probability where c and p are constants. This point c is a measure of location in the sense that c locates the $p\%$ point in the population.

3.42. Percentiles. The pth percentile point is that value d of a variate such that $P\{x \leqslant d\}=p\%=(0{\cdot}01)p$. where p and d are constants. For example the first percentile point, say p_1 is such that $P\{x \leqslant p_1\}=0{\cdot}01$. By this notation p_{10} is such that $P\{x \leqslant p_{10}\}=0{\cdot}1$. $p_{10}, p_{20}, ...$ are called the decile points. p_{25}, p_{50} and p_{75} are called the quartile points and they are also denoted by Q_1,

Q_2 and Q_3 respectively. That is, Q_2 is such that $P\{x<Q_2\} = P\{x>Q_2\}$. This is evidently a measure of central tendency of the distribution. Q_2 is also called the Median of the distribution.

Ex. 3.42.1. *Find the Median of the following probability distribution*

$$f(x)=\begin{cases} 2x \text{ for } 0<x<1 \\ 0 \text{ elsewhere} \end{cases}$$

Sol. By definition, if M is the median, then

$$P\{x<M\}=P\{x>M\}$$

But $$P\{x<M\}=F(M)$$

$$=\int_{-\infty}^{M} f(x)dx=\int_{0}^{M} 2x\, dx$$

$$=2\frac{M^2}{2}=M^2=\int_{M}^{1} 2x\, dx=1-M^2$$

$$\therefore \qquad M^2=1/2 \quad \text{or} \quad M=1/\sqrt{2}$$

(x is always positive here. Hence the negative root is not admissible).

Comments. It may be noticed that if we are given a set of numbers or observations then the median of this set may be defined as that number for which the number of observations less than it is equal to the number of observations greater than it. For example among the numbers 1, 7, 25 the number 7 is the median. In a probability density function, that is when X is continuous,

$$P\{x<M\}=\tfrac{1}{2}=P\{x>M\}.$$

Some idea about the quartiles is obtained from Fig. 3.12.

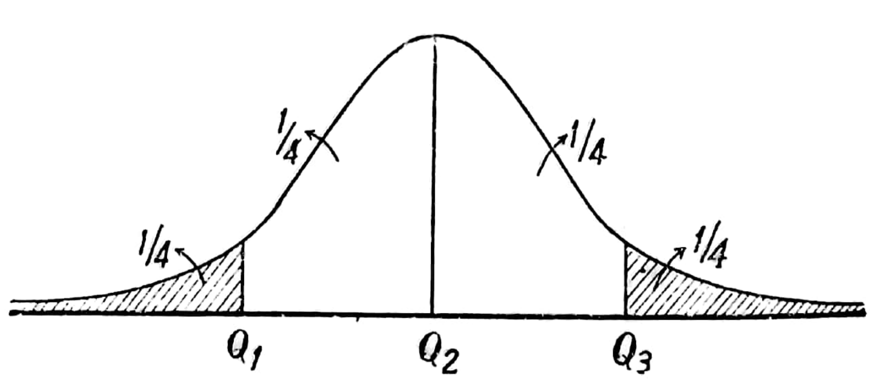


Fig. 3.12.

3.43. Measures of Central Tendency. It is seen that Q_2 is a measure of central tendency because $Q_2 = M =$ Median, is such that $P\{x < M\} = P\{x > M\}$ and when X is a continuous *s.v.* then

$$P\{x \leqslant M\} = \tfrac{1}{2} = P\{x \geqslant M\}.$$

Other measures of central tendency are the mean and the mode. $E(X) = \mu$, is called the mean or the mean value of the *s.v.*X. or of the population designated by X and is a good measure of central tendency. E(X) may be called the centre of gravity of the probability distribution. In a discrete distribution, if the *s.v.* takes the values $x_1, x_2, \ldots x_n$ with probabilities $p_1, p_2, \ldots p_n$ respectively where $\Sigma p_i = p_1 + p_2 + \ldots\ldots + p_n = 1$,

then
$$E(X) = \sum_{i=1}^{N} x_i p_i = \frac{\Sigma x_i p_i}{\Sigma p_i}$$

may be considered to be the centre of gravity of the system of masses $p_1, p_2, \ldots\ldots p_n$ at the points $x_1, x_2, \ldots\ldots, x_n$ respectively. (See Fig. 3.13).

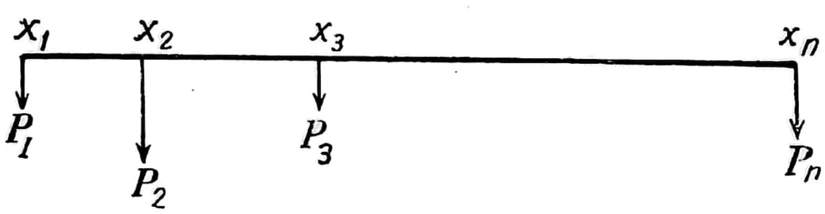


Fig. 3.13.

Another measure of central tendency is the mode. Mode may be defined as that value of the variate which occurs more frequently or that value of the variate corresponding to a maximum probability. For example if a discrete *s.v.* takes the values 0, 1, 2, 3 with probabilities 1/4, 1/2, 1/8, 1/8, then the mode may be taken

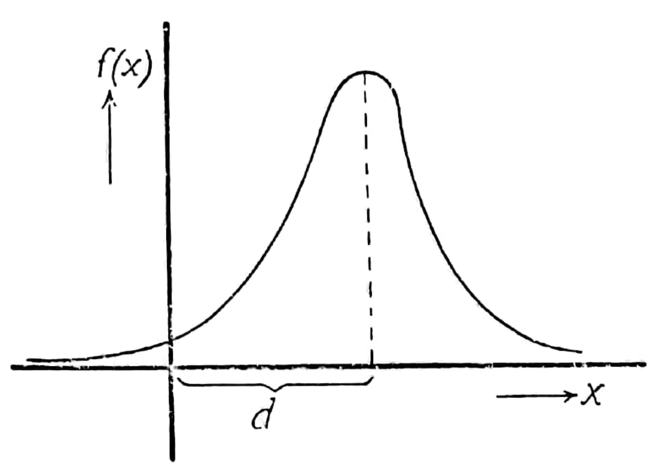


Fig. 3.14 (*a*)

as 1. The continuous distribution in Fig. 3.14 (*a*) has only one maximum point and is called unimodal. If a distribution has

more than one maximum point it is called multimodal. See Fig. 3.14 (*b*). In Fig. 3.14 (*b*), d_1 and d_2 are the modes, and the corresponding distribution is bimodal.

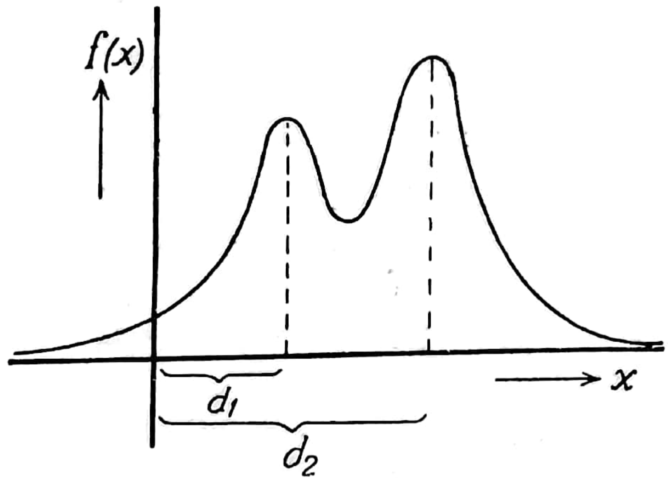


Fig. 3.14 (*b*).

Ex. 3.43.1. *Find the mode or modes of the following distribution* :

$$f(x)=\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\quad for\quad -\infty<x<\infty$$

Sol.
$$\frac{d}{dx}f(x)=-\frac{x}{\sqrt{2\pi}}\,e^{-x^2/2}$$
$$=0\Rightarrow x=0.$$

$\therefore$ The mode is 0. It may be noticed that $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ is a curve symmetric about the y-axis and the only maximum is at $x=0$.

Comments. The mode of a given set of numbers or observations may be defined as that number which occurs most frequently. For example in the following set of numbers 1, 2, 3, 3, 3, 4, 4, 5 the mode is 3.

3.44. Measures of Dispersion. A statistician may be interested in the scatter in a population. Scatter or dispersion may be broadly classified into three types : (1) the extent to which the elements are dispersed, that is, the range of the population, (2) the scatter among individual elements, (3) the scatter or dispersion of the elements from a point of reference. For example consider two townships A and B. Let the average income of the citizens in A be \$10,000 a year. If the average income of the citizens in B is also \$10,000 a year, from this information alone we cannot say that the two townships A and B are the same as far as the income distribution is concerned. In A there may be a few millionaires while the majority may be poor. In B perhaps everyone has an income around \$10,000 per year. If we have a measure

of the dispersion or scatter of the incomes from this average \$10,000 in A and B then we can say something more about the income distribution in A and B. Range and interquartile range are used to measure the extent to which the individuals are scattered. The Range is defined as the difference between the largest and smallest value that a discrete *s.v.* takes, with non-zero probability or it is the total range of a continuous *s.v.* where the density function is non-zero, For example if a stochastic variable takes the values $x_1, x_2 \dots\dots x_n$ where $x_1 \leqslant x_2 \leqslant \dots\dots \leqslant x_n$ then the Range is $x_n - x_1$. If a continuous *s.v.* is defined in the interval (0, 1) then the Range is $(1-0)=1$. $(Q_3-Q_1)/2$ is called the inter-quartile range. For measures of dispersion among individuals we use Gini's mean difference. The reader may refer to *Advanced Theory of Statistics* Vol. I by M.G Kendall and A. Stuart, for further information about the different measures.

The mean deviation from a point m may be taken as a measure of dispersion from the point m where m is a given quantity. This mean deviation measures the absolute values of the deviations from m, on the average. Similarly a root mean square deviation from m may also be taken as a measure of dispersion from m. In particular when $m=E(X)$ the root mean square deviation is the standard deviation. So standard deviation may be considered to be a measure of dispersion from $\mu=E(X)$. In general a pth root of the pth absolute moment about m may be taken as a measure of dispersion from m. It may be noticed that if $x=(x_1, \dots\dots, x_n)$ is a set of observations and if $d=(x_1-m, x_2-m, \dots, x_n-m)$, then a 'norm' of d say, $\| d \|$, with the condition that $\| d \| = 1$ if $| x_i - m | = 1$ for $i=1, 2, \dots, n$, may define a measure of dispersion or scatter from the point m in the set of observations $x_1, \dots, x_n$. Standard deviation mean deviation, roots of absolute moments are all special cases of this definition. The concept of 'norm' was introduced in chapter 1. For comparing two populations usually a measure called the coefficient of variation =(standard deviation)/(mean), if mean is non-zero and if the s.v's take positive values, is used.

Ex. 3.44.1. *Find the mean deviation and root mean square deviation from 4 for the following distribution.*

$$f(x)=\begin{cases} \dfrac{2x}{9} & \text{for } 0<x<3 \\ 0 & \text{elsewhere.} \end{cases}$$

Sol. The mean deviation from 4 is equal to $E\,|X-4|$

$$= \int_{-\infty}^{\infty} |x-4| \, f(x) \, dx$$

$$=\int_0^3 |x-4| \left(\frac{2x}{9}\right) dx$$

$$=\int_0^3 (4-x) \frac{2}{9} x\, dx$$

(since $x<4$, $4-x$ is always positive)

$$=\frac{2}{9}\left[4\frac{x^2}{2}-\frac{x^3}{3}\right]_0^3$$

$$=2.$$

The root mean square deviation from 4 is

$$=\{E(X-4)^2\}^{\frac{1}{2}}$$

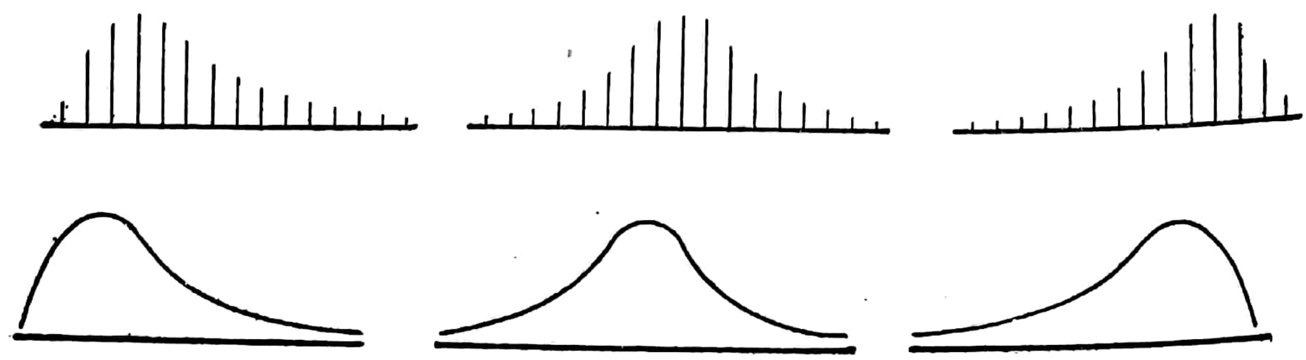
But $$E(X-4)^2=\int_0^3 (x-4)^2 \frac{2}{9} x\, dx$$

$$=\frac{2}{9}\left[\frac{x^4}{4}-8\frac{x^3}{3}+8x^2\right]_0^3=\frac{9}{2}$$

$\therefore \quad \{E(X-4)^2\}^{\frac{1}{2}}=\sqrt{9/2}=3/\sqrt{2}.$

Comments. Similarly other measures of dispersion from the point 4 or from E(X) or from any other point may be evaluated for this population designated by the s.v., X with the probability function as defined in this example.

3.45. Measures of Skewness. If a probability distribution is symmetric then it is easily seen that the odd central



Skewed to the right (a) *Symmetric* (b) *Skewed to the left* (c)

Fig. 3.15.

moments $\mu_1, \mu_3, \mu_5 \ldots$ are all zero. So the odd central moments may be taken as a measure of skewness or lack of symmetry in the probability distribution. To be independent of the units of measurements, μ_3/σ^3, $\mu_5/\sigma^5, \ldots$ may be taken as measures of skewness where σ is the standard deviation ($\sigma \neq 0$ is assumed). If an odd central moment is zero this does not necessarily mean that the distribution is symmetric. So these different measures are measures of skewness only to some extent. Different types of skewed distributions are shown in Fig. 3.15.

Ex. 3.45.1. *Examine the symmetry and evaluate the coefficient of variation in the following probability distribution :*

$$f(x) = \begin{cases} 1/4 & \text{for } x=2 \\ 1/2 & \text{for } x=3 \\ 1/4 & \text{for } x=4 \\ 0 & \text{elsewhere} \end{cases}$$

Sol. $$E(X) = 2\times\tfrac{1}{4} + 3\times\tfrac{1}{2} + 4\times\tfrac{1}{4} = 3$$

$\therefore$ $$\mu_3 = (2-3)\times\tfrac{1}{4} + (3-3)\times\tfrac{1}{2} + (4-3)\times\tfrac{1}{4} = 0.$$

Similarly all the odd moments $\mu_5, \mu_7, \ldots = 0$

$\therefore$ The distribution may be considered to be symmetric

$$\mu_2 = E(X-3)^2 = (2-3)^2\times\tfrac{1}{4} + (3-3)^2\times\tfrac{1}{2} + (4-3)^2\times\tfrac{1}{4}$$
$$= 1/2$$

$\therefore$ $$\sigma = \sqrt{\mu_2} = 1\sqrt{2}.$$

$\therefore$ The coefficient of variation

$$= \frac{\sigma}{\mu} = \frac{1}{\sqrt{2}\times 3}$$

$$= \frac{1}{3\sqrt{2}}.$$

Comments. For continuous distributions symmetry may be checked in a similar fashion.

3.46. Kurtosis. Kurtosis or peakedness of a probability distribution is usually measured by

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3.$$

(where γ is a greek letter called gamma)

For a Gaussian or Normal distribution $\mu_4/\sigma^4 = 3$. This distribution is taken as a standard to measure Kurtosis. Distributions for

which $\gamma_2=0, >0, <0$ are called mesokurtic, leptokurtic and platykurtic, respectively.

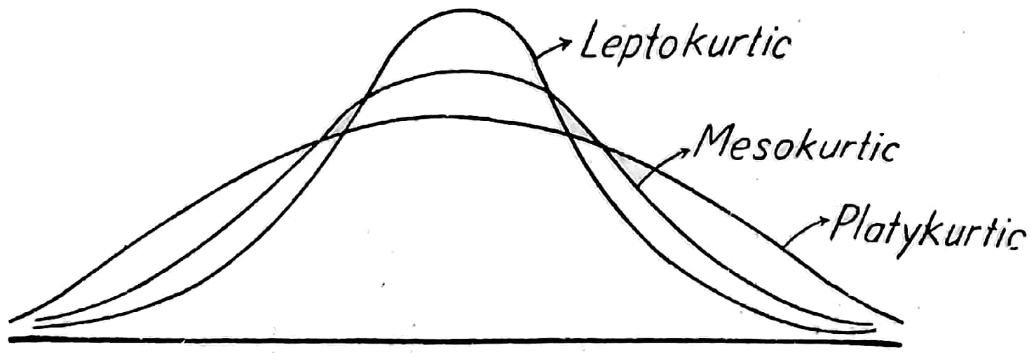


Fig. 3·16.

However if $\gamma_2=, >0, <0$ the shapes of the curves need not be as shown in Fig. 3.14. So γ_2 does not tell much about the shape of the distribution.

Exercises

3.26. Give one example each of a probability distribution which is (1) symmetric, (2) skewed to the right, (3) skewed to the left.

3.27. Evaluate, (1) Median, (2) Second decile point, (3) the centre of gravity$=E(X)$. (4) the range, (5) the interquartile range, (6) the standard deviation, of the following distribution.

$$f(x)=\begin{cases} 1/\theta \text{ for } 0<x<0 \text{ and } \theta>0 \text{ is a parameter.} \\ 0 \text{ elsewhere} \end{cases}$$

3.28. The following are some important inequalities, given for the information of the reader. The reader may try to prove them

(1) $E\,|\,X\,| \leqslant \{E\,|\,X\,|^r\}^{1/r}$ for $r \geqslant 1$;

(2) $\{E\,|\,X\,|^r\}^{1/r} \leqslant \{E\,|\,X\,|^s\}^{1/s}$ for $0<r\leqslant s$;

(3) $\log E\,|\,X\,|^r$ is a convex function of r. (A function $h(x)$ is convex if $h(\alpha x+\beta y) \leqslant \alpha h(x)+\beta h(y)$ for all x, y, $\alpha>0$, $\beta>0$, $\alpha+\beta=1$) ;

(4) If $h(X)$ is a convex function of X and if $E(X)$ exists then $h[E(X)] \leqslant E[h(X)]$. (This is known as Jensen's inequality).

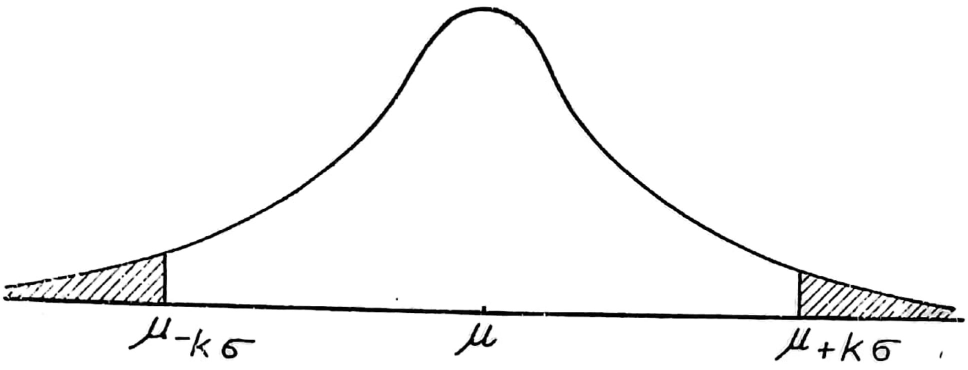
3.5. CHEBYSHEV'S THEOREM

In section 3.44 we had seen that the standard deviation may be taken as a measure of dispersion from the expected value. Now we will prove a theorem due to a Russian Mathematician Chebyshev (also spelled as Tchebycheff) which will throw some light on the importance of the standard deviation in statistical analysis. The theorem states that

$$P\{\,|\,x-\mu\,|>k\sigma\} < \frac{1}{k^2} \tag{3·39}$$

where k is an arbitrary positive constant, $\mu=E(X)$ and σ is the standard deviation, *i.e.*, the probability that the absolute value of $x-\mu$ is greater than k times the standard deviation is less than $1/k^2$ or in other words the probability that x is less than $\mu-k\sigma$ or

greater than $\mu+k\sigma$, is less than $1/k^2$. This gives the probability of closeness of a *s.v.*, X to its expected value.



Fig, 3.17.

According to the theorem the probability that x falls below $\mu-k\sigma$ or above $\mu+k\sigma$, is less than $1/k^2$ or the shaded area in Fig. 3.17 is less than $1/k^2$.

i.e., $$P\{\,|x-\mu|>k\sigma\}=\int_{-\infty}^{\mu-k\sigma} f(x)dx+\int_{\mu+k\sigma}^{\infty} f(x)dx<\frac{1}{k^2}$$

if X is continuous

Proof. Let X be a continuous *s.v.*

$$\sigma^2=E[X-E(X)]^2=\int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx$$

$$=\int_{-\infty}^{\mu-k\sigma}(x-\mu)^2 f(x)dx+\int_{\mu-k\sigma}^{\mu+k\sigma}(x-\mu)^2 f(x)dx$$

$$+\int_{\mu+k\sigma}^{\infty}(x-\mu)^2 f(x)dx \qquad (3\cdot40)$$

$$\therefore \quad \sigma^2 \geqslant \int_{-\infty}^{\mu-k\sigma}(x-\mu)^2 f(x)dx+\int_{\mu+k\sigma}^{\infty}(x-\mu)^2 f(x)dx \qquad (3\cdot41)$$

But in the intervals $(-\infty,\ \mu-k\sigma)$ and $(\mu+k\sigma,\ \infty)$,

$$|x-\mu|>k\sigma \Rightarrow (x-\mu)^2>k^2\sigma^2$$

$$\therefore \quad \sigma^2 > \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(x)dx \tag{3.42}$$

$$\Rightarrow \quad \frac{1}{k^2} > \int_{-\infty}^{\mu-k\sigma} f(x)dx + \int_{\mu+k\sigma}^{\infty} f(x)dx = \mathrm{P}\{\,|\,x-\mu\,|>k\sigma\}$$

The proof when X is discrete is left to the reader.

For a more rigorous statement and generalizations of the theorem the reader may refer to Mathematical Methods of Statistics by H. Cramer and Linear Statistical Inference and Its Applications by C.R. Rao. This inequality is very important due to many reasons. The proximity of a *s.v.* to its expected value, measured in terms of a measure of dispersion, namely, the standard deviation, is given by this inequality. The exact value of $\mathrm{P}\{\,|\,x-\mu\,|>k\sigma\}$ can be evaluated by knowing the distribution of X. But here no emphasis is put on the functional form of the probability function of X. Hence this inequality may be considered to be a 'distribution free' property, in the sense that the property does not depend on the distribution of X. Instead of using the standard deviation we can use any other measure of dispersion and obtain the corresponding inequalities.

Ex. 3.5.1. *The probability of survival in case of a particular disease D is found to be 0·80. One hundred people are attacked by D in a particular area. If X denotes the number of survivals, assuming that X follows a Binomial distribution with parameters N=100 and p=0·80, find (1) an upper bound for the probability that the number of survivals will be either less than 68 or greater than 92; (2) a lower bound for the probability that the number of survivals is between 86 and 92.*

Sol. (1) When X follows Binomial distribution,

$$f(x)=\binom{\mathrm{N}}{x} p^x\,(1-p)^{\mathrm{N}-x}$$

for $x=0, 1, \ldots\ldots, \mathrm{N}$, it can be seen from Ex. 3.35.1 and Ex. 3.35.2 that $\mathrm{E(X)=N}p$ and variance of $\mathrm{X=N}p(1-p)$ and hence the standard deviation

$$\sigma=[\mathrm{N}p(1-p)]^{\frac{1}{2}}$$

Here $\quad \mathrm{N}=100,\ p=0.80$

and $\quad 1-p=0.20.$

Therefore $\quad \mathrm{N}p=\mu=100\times0.80=80$

and $\quad \sigma^2=\mathrm{N}p(1-p)=100\times0.80\times0.20=16$

or $\quad \sigma=4.$

But $\qquad 80+3\times 4=92$ and $80-3\times 4=68,$

that is, $\qquad \mu+k\sigma=92$ and $\mu-k\sigma=68$ where $k=3.$

According to Chebyshev's inequality,

$$P\{\,|\,x-\mu\,|\,>k\sigma\}<1/k^2,$$

that is, $\qquad P\{\,|\,x-80\,|\,>12\}<1/9,$

or $\qquad P\{x<68 \text{ or } x>92\}<1/9.$

(In this example, since the population is known to be Binomial, we can calculate the exact probabilities instead of the approximate probabilities given by the inequalities, if we have a table of Binomial probabilities).

Hence an upper bound for the required probability is 1/9.

(2) $\qquad P\{68\leqslant x\leqslant 92\}=1-P\{x<68 \text{ or } x>92\}\geqslant 1-\dfrac{1}{k^2}$

where $\qquad k=3.$ $\qquad$ (3·44)

$\therefore$ The required probability limit

$$=1-1/9=8/9.$$

Comments. It may be noticed that

$$P\{\,|\,x-\mu\,|\,<k\sigma\}>1-\frac{1}{k^2}.$$

This may be easily derived from Chebyshev's inequality. For $k=1, 2, 3,\ldots$ we get the various probability statements about the closeness of X to μ in terms of σ.

Ex. 3.5.2. *In Ex. 3.5.4 it is given that* $P\{\,|\,x-\mu\,|\,>k\}<0{\cdot}25$ *Find k?*

Sol. By Chebyshev's theorem

$$P\{\,|\,x-\mu\,|\,>k\sigma\}<\frac{1}{k^2}.$$

If k is replaced by $\dfrac{k}{\sigma}$ then the inequality may be written as

$$P\{\,|\,x-\mu\,|\,>k\}<\frac{\sigma^2}{k^2}. \qquad (3{\cdot}45)$$

We are given that $P\{\,|\,x-\mu\,|\,>k\}<0.25.$

$\therefore$ k may be obtained by taking

$$\frac{\sigma^2}{k^2}=0.25$$

or $$k=\frac{\sigma}{0{\cdot}5}=2\sigma=8.$$

Comments. It may be noticed that Chebyshev's inequality as stated in the above theorem does not give the least upper bound for the probability that $|x-\mu| > k\sigma$. Still smaller upper bounds for $P\{|x-\mu| > k\sigma\}$ may be established. This will not be discussed here.

Exercises

3.29. If X follows a Binomial distribution

$$f(x, \theta)=\binom{N}{x} p^x q^{N-x} \text{ where } q=1-p,\ p=0.4,\ N=10,$$

find k for the following problems.

(1) $P\{|x-\mu| > k\sigma\} < 0.25$,

(2) $P\{|x-\mu| > k\} < 0.5$, where $\mu=E(X)$ and σ is the standard deviation of X.

3.30. If X follows a Binomial distribution with the parameters N and p specified as $N=40$ and $p=1/2$, obtain lower limits for the probabilities,

(1) $P\{|x-\mu| < 10\}$,

(2) $P\{|x-\mu| < 20\}$,

(3) $P\{|x-\mu| < 30\}$,

by using Chebyshev's theorem and also explain these limits in words.

3.31. A Gamma distribution is defined as,

$$f(x)=x^{\alpha-1}\, e^{-x}/\Gamma(\alpha) \text{ for } x>0,\ \alpha>0 \text{ and } f(x)=0$$

elsewhere, where $\Gamma(\alpha)$ (gamma alpha) is defined as

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}\, e^{-x}\, dx.$$

It can be seen that $\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$ and $\Gamma(1/2)=\sqrt{\pi}$. Obtain inequalities for the following probabilities by using Chebyshev's theorem.

(1) $P\{|x-\mu| < 2\sigma\}$,

(2) $P\{|x-\mu| > 2\sigma\}$,

(3) $P\{|x-\mu| > 2\}$, where $\mu=E(X)$ and σ=the standard deviation of x.

3.32. Prove the following inequalities :

(1) $P\{|x-\mu| \geqslant k\,\beta_r^{1/r}\} \leqslant 1/k^r$ for $r \geqslant 1$;

(2) $P\{|x-\mu| \geqslant k\} \leqslant \beta_r/k^r$ for $r \geqslant 1$, where $\beta_r=E\,|X-\mu|^r$.

[**Hint.** Proceed in a similar fashion as in the proof of Chebyshev's inequality].

3.6. COMPLETENESS

We denoted a probability function by $f(x, \theta)$ where θ stands for all the parameters in the probability function. We have also seen that if the parameters are specified then the probability

function is completely specified. For example, in an exponential distribution with one parameter the density function is,

$$(x, \theta)=\begin{cases}\left(\frac{1}{\theta}\right) e^{-x/\theta} & 0<x<\infty, \theta>0 \\ 0 \text{ elsewhere.} \end{cases}$$

For $\theta=10$ and $\theta=24.2$ we get two exponential distributions with no parameters (or completely specified). In order that $f(x, \theta)$ be a probability function the only restriction on θ is that $\theta>0$. In other words θ is a constant which is defined as $0<\theta<\infty$. For the various possible values of θ, $f(x, \theta)$ defines a family of exponential distributions. In general if a probability function has a parameter it will define a family of probability distributions, namely, all the distributions with the probability functions having the same functional form except for the value of the parameter.

In the statistical theory of estimation and related problems we often need a function of the *s.v.* X. say $\psi(X)$, such that E $\psi(X)$ $=\theta$, where θ is a parameter. If there exist two functions such that the expected values of both the functions are θ then $\psi(X)$ is not unique. Simplifying this problem and putting this in a simple from we would like to ask the question, does there exist a non-zero function, say, $\phi(x)$ such that $E\phi(X)=0$? Since $E(0)=0$, if there exists a non-zero $\phi(x)$ such that $E\phi(X)=0$ then $\phi(x)$ is not unique or the probability function of the *s.v.*, X has the property that $E\phi(X)=0 \not=> \phi(x)=0$ for all x and for possible values of the parameter in the probability function of X.

Definition. Let $f(x, \theta)$ be a family of probability functions. Let $\phi(x)$ be a continuous function of x, independent of the parameter θ. If $E\phi(X)=0=>\phi(x)=0$ for all x and for all possible values of the parameter θ (except for a set with probability zero) then $f(x, \theta)$ is said to be a complete family of probability functions. This implies that for a complete family of probability functions, there exists no non-zero continuous function $\phi(x)$ such that $E\phi(X)=0$ for all x and for all θ.

Ex. 3.6.1. *Check whether the following probability measures are complete :*

(*a*) $f(x)=(2\pi)^{-1/2} \quad e^{-x^2/2}, \qquad -\infty<x<\infty.$

(*b*) $f(x, \beta)=(2\pi\beta^2)^{-1/2} \quad e^{-x^2/2,\beta^2} -\infty<x<\infty,\ \beta>0.$

(*c*) $f(x, \alpha)=(2\pi)^{-1/2} \quad e^{-(x-\alpha)^2/2}, -\infty<x<\infty,\ -\infty<\alpha<\infty.$

Sol. (*a*) Consider the function $\phi(x)=x$,

$$E(X)=\int_{-\infty}^{\infty} (2\pi)^{-1/2} x\, e^{-x^2/2}\, dx$$

$$= (2\pi)^{-1/2} \int_{-\infty}^{\infty} x \, e^{-x^2/2} \, dx = 0.$$

Evidently x is a non-zero continuous function of x and hence $f(x)$ is not complete.

$$(b) \; E(X) = (2\pi\beta^2)^{-1/2} \int_{-\infty}^{\infty} x \, e^{-x^2/2\beta^2} \, dx = 0.$$

Irrespective of the value of β there exists a non-zero function independent of β, namely $\phi(x) = x$, such that $E\phi(X) = 0$ and hence $f(x, \beta)$ is not complete.

(c) It can be shown that

$$(2\pi)^{-1/2} \int_{-\infty}^{\infty} \phi(x) \, e^{-(x-\alpha)^2/2} \, dx$$

$$= 0 \Rightarrow \phi(x) = 0,$$

almost everywhere. The proof is beyond the scope of this book. (One method of proving this result is by using the uniqueness of Laplace transforms.)

Exercise

3.33. Construct 2 examples of a family of probability function which is not complete.

BIBLIOGRAPHY

1. H. **Cramer**. Mathematical Methods of Statistics. Princeton University Press 1961.

2. D.A.S. **Fraser**. Statistics, An Introduction. John Wiley & Sons Inc. New York, 1958.

3. D.A.S. **Fraser**. Non Parametric Methods in Statistics. John Wiley, N.Y. 1957.

4. J.E. **Freund**, Mathematical Statistics. Prentice-Hall, N.J. 1964

5. P.G. **Hoel**, Introduction to Methematical Statistics, John Wiley N.Y. 1964.

6. M.G. **Kendall** & A. **Stuart**. The Advanced Theory of Statistics, Vol. I, Charles Griffin Co. London, 1958.

7. A.M. **Mood** & F.A. **Graybill**, Introduction to the Theory of Statistics, McGraw-Hill, Book Co. Inc. N.Y, 1963.

8. M.H. **Protter** & C.B. **Morrey** Jr. College Calculus with Analytical Geometry. Addison Wesley, Inc. Reading, Mass. 1965.

9. C.R. **Rao**. Linear Statistical Inference and Its Applications. John Wiley & Sons, N.Y. 1965.

10. G.P. **Wadsworth** & J.G. **Bryan**. Introduction to Probability and Randon Variables. McGraw-Hill, N.Y. 1960.

CHAPTER 4

SPECIAL UNIVARIATE DISTRIBUTIONS

4.0. Introduction. Here some of the most commonly used univariate (one variate) probability distributions will be discussed. In chapter 3 we defined probability distributions and a general notation $f(x, \theta)$ was introduced where θ denotes all the parameters in the probability function. For convenience some of the important univariate discrete and univariate continuous distributions are given in tabular forms. In the later sections these distributions and some of their important properties are individually discussed.

4.1. DISCRETE AND CONTINUOUS PROBABILITY MODELS

In practical situations where the behaviour of a particular characteristic (say, one stochastic variable) is under study, we may not know the appropriate probability distribution for the situation under consideration. By examining the experimental conditions, such as the possible outcomes of an experimental trial, the probability of an outcome, independence etc., we may be able to set up an appropriate probability model. Once we have a probability model or a probability distribution the experimental results can be studied in greater detail. In sections 4·11 and 4·12 some of the most frequently used univariate probability models are given and the special experimental conditions for which these models are appropriate, are also discussed in later sections.

4.11. Discrete distributions.

Name	*Probability function, $f(x, \theta)$*	*Parameters θ*
1. The Binomial distribution.	$\binom{N}{x} p^x(1-p)^{N-x}$ for $x=0, 1, \ldots\ldots, N$ and $f(x, \theta)=0$ elsewhere.	(N, p) $0<p<1$ N-positive integer.
2. The Hypergeometric distribution.	$\dfrac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}$ for $x=0, 1, \ldots\ldots, n$. and 0 elsewhere.	(a, b, c) all positive integers.

Name	*Probability function $f(x, \theta)$*	*Parameters θ*
3. The Poisson distribution.	$\lambda^x e^{-\lambda}/x\,!$ for $x=0, 1, \ldots\ldots \infty$ and 0 elsewhere	(λ) $\lambda>0$
4. The Negative Binomial distribution.	$\binom{x+k-1}{x} p^k(1-p)^x$ for $x=0, 1, \ldots\ldots, \infty$ and 0 elsewhere.	(p, k) $0<p<1$ k-positive integer.
,,	$\binom{x-1}{k-1} p^k(1-p)^{x-k}$ for $x=k, k+1, \ldots\ldots \infty$, and 0 elsewhere.	,,
5. The discrete uniform distribution.	$1/n$ for x equals some $x_1, x_2, \ldots\ldots, x_n$ and 0 elsewhere.	(n) n-positive integer.
6. The discrete Geometric distribution.	$\theta(1-\theta)^{x-1}$ for $x=1, 2, \ldots\ldots, \infty$ and 0 elsewhere.	(θ) $0<\theta<1$.

4.12. Continuous Distributions.

Name	*Probability density $f(x, \theta)$*	*Parameters θ*
1. The Uniform or rectangular distribution.	$f(x, \theta)=\dfrac{1}{\beta-\alpha}$ for $\alpha<x<\beta$ $=0$ elsewhere.	(α, β) $\alpha>0, \beta>0$
,,	$f(x, \theta)=\dfrac{1}{\theta}$ for $0<x<\theta$, $=0$ elsewhere.	(θ) $\theta>0$
2. The Exponential distribution.	$f(x, \theta)=\dfrac{1}{\theta}\, e^{-\frac{x}{\theta}}$ for $0<x<\infty$, $=0$ elsewhere.	(θ) $\theta>0$
***3.** The Gamma distribution.	$f(x, \theta)=\dfrac{1}{\beta^{\alpha}\,\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-x/\beta}$, for $0<x<\infty$, $=0$ elsewhere.	(α, β) $\alpha, \beta>0$
,,	$f(x, \theta)=\dfrac{1}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-x}$ for $0<x<\infty$ $=0$ elsewhere	(α) $\alpha>0$

*$\Gamma(\alpha)$ (Gamma alpha) is defined as

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}\, e^{-x}\, dx$$

where Γ is the Greek capital letter Gamma. It can be proved that $\Gamma(\alpha)=(\alpha-1).\,\Gamma(\alpha-1)$ and $\Gamma(1/2)=\sqrt{\pi}$. If α is a positive integer $\Gamma(\alpha)=(\alpha-1)\,!$

Name	*Probability density* $f(x, \theta)$	*Parameters* θ
*4. The Beta distribution.	$f(x, \theta)=\frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0<x<1$ $=0$ elsewhere	(α, β) $\alpha, \beta>0$
**5. The Cauchy distribution.	$f(x, \theta)=\frac{1}{\pi[1+(x-\theta)^2]}, \quad -\infty<x<\infty$	(θ) $-\infty<\theta<\infty$
,,	$f(x)=\frac{1}{\pi[1+(x^2-\theta)]}, \quad -\infty<x<\infty$	(.)
6. The Gaussian or normal distribution.	$f(x, \theta)=\frac{1}{\beta\sqrt{2\pi}} e^{-\frac{(x-\alpha)^2}{2\beta^2}}, \quad -\infty<x<\infty$	(α, β) $-\infty<\alpha<\infty$, $\beta>0$
,,	$f(x, \theta)=\frac{1}{\beta\sqrt{2\pi}} e^{-\frac{x^2}{2\beta^2}}, \quad -\infty<x<\infty$	(β) $\beta>0$
,,	$f(x, \theta)=\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\alpha)^2}{2}}, \quad -\infty<x<\infty$	(α) $-\infty<\alpha<\infty$
,,	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty<x<\infty$	(.)
7. The Pearson curves.	$\frac{d}{dx} f(x, \theta)=f(x, \theta) \frac{d-x}{a+bx+cx^2}$	(a, b, c, d)
8. The student's *t* distribution.	$f(x, \theta)=\frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\ \Gamma\left(\frac{k}{2}\right)}\left(1+\frac{x^2}{k}\right)^{-\left(\frac{k+1}{2}\right)}$ for $-\infty<x<\infty$	(k) k-positive integer

*$B(\alpha, \beta)$ (beta alpha beta) is defined as

$$B(\alpha, \beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx,$$

where B is the Greek capital letter beta and β is the Greek small letter beta. It can be proved that

$$B(\alpha, \beta)=\frac{\Gamma(\alpha)\,.\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

**(.) means that there is no parameter in this probability distribution.

Name	*Probability density* $f(x, \theta)$	*Parameters* θ
9. The Chi-square distribution	$f(x, \theta) = \dfrac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} x^{(k/2)-1} e^{-x/2}$ for $0 < x < \infty$; $= 0$ elsewhere	(k) k-positive integer
10. The F-distributive	$f(x, \theta) = \dfrac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{2}\right)^{m/2} \dfrac{x^{(m/2)-1}}{\left(1+\frac{m}{n}\right)^{\frac{m+n}{2}}}$ $= 0$ elsewhere.	(m, n) m, n positive integers.

The Pearsonian system of curves will give a number of probability distributions for various values of a, b, c and d. This may be seen from an exercise given at the end of section 4.8. The differential equation in 7 gives the density functions. The range of x may be determined by the axioms for a probability function. The reader is advised to learn by heart the parts involving x in the various density functions together with the range of the stochastic variables. The constants in the various density functions (the parts independent of x) may be determined by the property that the total measure is unity,

i.e., $$\int_{-\infty}^{\infty} f(x, \theta)dx = 1.$$

or $$\sum_{-\infty < x < \infty} f(x, \theta) = 1.$$

4.2. THE BINOMIAL DISTRIBUTION

This is the most widely used univariate discrete distribution. The probability function is

$$f(x, \theta) = \binom{N}{x} p^x q^{N-x} \text{ where } q = 1 - p$$

and N and p are parameters, *i.e.*, for a given Binomial distribution N and p are constants. For different values of N and p different Binomial distributions are obtained. So a Binomial distribution is completely known if N and p are given.

This distribution may be easily obtained by considering an experiment of repeated trials. Suppose that we are interested in finding out the probabilities of getting exactly 3 heads in 10 trials

of tossing a balanced coin, getting 5 boys among 20 babies born in a hospital given that the probability that a new born baby in that hospital is a boy is 1/3, four out of 50 pneumonia patients dying given that the probability of death in case of pneumonia is 0.01 etc. Binomial probability law may be applied in situations where (1) a trial results in either a success or a failure, (2) the probability of occurrence of an event (we usually call this the probability of a success in the sense, favourable to the event under consideration. For example if we are interested in the event of a death then a success is a death etc.) remains the same from trial to trial, (3) the trials are independent. If these three conditions are satisfied then the probability of getting exactly x successes in N trials may be seen to be $\binom{N}{x} p^x (1-p)^{N-x}$ where p is the probability of a success in any trial. Suppose that the first x trials are successes and the remaining $N-x$ are failures, then the probability of getting the first x successes and the remaining $N-x$ failures is $p^x(1-p)^{N-x}$. (Since all the trials are independent the probability is the product of individual probabilities). Suppose that the first trial is a failure, the next x are successes, and the remaining ones are failures. The probability for this is $q\, p^x.\, q^{N-x-1} = p^x \; q^{N-x}$ where $q=1-p$. So if we specify any subset x of trials resulting in successes and the remaining trials resulting in failures the probability for this is $p^x \; q^{N-x}$. But a subset of x elements from a set of N elements may be selected in $\binom{N}{x}$ ways. Therefore the probability of getting exactly x successes in N trials is $\binom{N}{x} p^x \; q^{N-x}$.

It is seen that the Binomial probability law applies in situations where,

(1) Any trial results in a success or a failure ;

(2) There are N repeated trials which are independent ;

(3) The probability of a success in any trial is p.

Such a situation may be called a Binomial probability situation. If we have a situation different from this then the Binomial law is not applicable. Some other probability distribution may be found out. So in a Binomial situation it is seen that,

$$f(x, \theta) = \begin{cases} \binom{N}{x} p^x \; q^{N-x} \text{ for } x = 0,\ 1,\ 2, \ldots N \\ 0 \text{ elsewhere.} \qquad 0 < p < 1,\ q = 1-p \end{cases} \tag{4.1}$$

This is called a Binomial distribution because the probabilities of getting 0, 1, 2, 3,... successes in N trials may be obtained as the first, second, third,... terms of the Binomial expansion

$$(q+p)^{N}=\binom{N}{0}p^{0}\ q^{N-0}+\binom{N}{1}p^{1}q^{N-1}+\ldots$$
$$+\binom{\ }{x}p^{x}\ q^{N-x}+\ldots\ldots+\binom{N}{N}p^{N}q^{0}.$$

But $q=1-p$ and therefore $q+p=1$. (4.2)

$$\therefore \quad 1=\binom{N}{0}p^{0}\ q^{N-0}+\ldots\ldots+\binom{N}{x}p^{x}\ q^{N-x}+\ldots$$
$$+\binom{N}{N}p^{N}q^{0}. \qquad (4.3)$$

i.e., the probabilities of getting 0, 1, 2,... Successes add up to unity. This may be expected because the event of getting 0 or 1 or 2 or...... or N successes is a sure event. It may be noticed that the Binomial coefficients are symmetric in the sense

$$\binom{N}{N}=\binom{N}{0};\ \binom{N}{1}=\binom{N}{N-1}, \text{ etc.}$$

But the Binomial probabilities or the different terms of the Binomial expansion of $(q+p)^{N}$ need not be symmetric. They are symmetric if $p=q=1/2$. If $p\neq q$ then

$$\binom{N}{0}p^{0}\ q^{N-0}\neq\binom{N}{N}p^{N}q^{0};\ \binom{N}{1}p^{1}\ q^{N-1}\neq\binom{N}{N-1}p^{N-1}q^{1} \text{ etc.}$$

Ex. 4.2.1. *Find the probability of getting (1) exactly 3 heads, (2) at least 3 heads, (3) at most 3 heads, if a coin is tossed five times, assuming that the probability of getting a head in any trial is 1/3.*

Sol. This is a Binomial situation of repeated independent trials. According to our notation $N=5$, $p=1/3$.

In (1) $x=3$, in (2) $x=3$ or 4 or 5 and in (3) $x=0$ or 1 or 2 or 3. Therefore the required probabilities are

$$(1)\quad \binom{5}{3}\left(\frac{1}{3}\right)^{3}\left(\frac{2}{3}\right)^{2}=40/243.$$

$$(2)\quad \binom{5}{3}\left(\frac{1}{3}\right)^{3}\left(\frac{2}{3}\right)^{2}+\binom{5}{4}\left(\frac{1}{3}\right)^{4}\left(\frac{2}{3}\right)+\binom{5}{5}\left(\frac{1}{3}\right)^{5}\left(\frac{2}{3}\right)^{0}$$
$$=\frac{51}{243}$$

$$(3)\quad \binom{5}{0}\left(\frac{1}{3}\right)^{0}\left(\frac{2}{3}\right)^{5}+\binom{5}{1}\left(\frac{1}{3}\right)^{1}\left(\frac{2}{3}\right)^{4}+\binom{5}{2}\left(\frac{1}{3}\right)^{2}\left(\frac{2}{3}\right)^{3}$$
$$+\binom{5}{3}\left(\frac{1}{3}\right)^{3}\left(\frac{2}{3}\right)^{2}.$$

$=1-$[Probability of getting at least 4 heads].

$$=1-\left[\binom{5}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^1+\binom{5}{5}\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)^0\right]=\frac{232}{243}.$$

Comments. Here the parameters are N and p which are specified as $N=5$ and $p=1/3$. This Binomial distribution is completely specified. It may be noticed that in a Binomial probability distribution if the various probabilities of getting 0, 1, 2,...... successes are represented by a histogram we get a symmetric histogram if $p=q=1/2$. Otherwise the histogram will be skewed to the right or to the left depending upon p. In this example the various terms may be evaluated by the help of a Binomial probability table. The Binomial probabilities for various values of N and p are tabulated. An extract of the table is given at the end of this book.

4.21. Moments. By definition

$$E(X)=\sum_{x=0}^{N} x\,.\binom{N}{x} p^x\; q^{N-x}$$

(when $x=0$ the corresponding term is zero.)

$$=\sum_{x=1}^{N} x\frac{N\,!}{x\,!\,(N-x)\,!}\, p^x\; q^{N-x}$$

$$=\sum_{x=1}^{N} \frac{N\,!}{(x-1)\,!\,(N-x)\,!}\, p^x\; q^{N-x}$$

$$=N.p\sum_{x=1}^{N}\frac{(N-1)\,!}{(x-1)\,!\,[(N-1)-(x-1)]\,!}\,p^{x-1}.\,q^{N-1-(x-1)}.$$

Put $x-1=y$ and $N-1=n$

$$E(X)=Np.\sum_{y=0}^{n}\frac{n\,!}{y\,!\,(n-y)\,!}\,p^y\; q^{n-y}$$

$$=Np\sum_{y=0}^{n}\binom{n}{y}p^y\; q^{n-y}$$

$Np(q+p)^n$ [Since $(q+p)^n$ when expanded gives all the terms in $\sum_{y=0}^{n}\binom{}{y}p^y\; q^{n-y}$]

$$=Np \text{ (since } q+p=1) \qquad (4.4)$$

$\therefore$ $E(X)=Np=\mu$ (say)

$$\text{Var }(X)=E[X-E(X)]^2=E(X^2)-\mu^2$$

$$E(X^2)=\sum_{x=0}^{N} x^2\frac{N\,!}{x\,!\,(N-x)\,!}\,p^x\; q^{N-x}$$

$$= \sum_{x=0}^{N} x(x-1) \frac{N!}{x!\,(N-x)!} p^x \; q^{N-x}$$

$$+ \sum_{x=0}^{N} x \frac{N!}{x(N-x)!} p^x \; q^{N-x}$$

[Since $x^2 = x(x-1) + x$]

$$= \sum_{x=2}^{N} x(x-1) \frac{N!}{x!\,(N-x)!} p^x q^{N-x} + \mu$$

$$= N(N-1)p^2 \sum_{x=2}^{N} \frac{(N-2)!}{(x-2)!\,(N-x)!} p^{x-2} \; q^{N-x} + \mu.$$

Put $N-2=n$, and $x-2=y$. Then the above equation is

$$= N(N-1)p^2 \sum_{y=0}^{n} \binom{n}{y} p^y \; q^{n-y} + \mu$$

$$= N(N-1)p^2 (q+p)^n + \mu$$

$$= N(N-1)p^2 + \mu$$

$$\therefore \quad E(X^2) = N(N-1)p^2 + Np \tag{4.5}$$

$$\therefore \quad \text{Var}\,(X) = E(X^2) - \mu^2 = N(N-1)p^2 + Np - (Np)^2$$

$$= Np - Np^2 = Np(1-p)$$

$$= Npq \tag{4.6}$$

$\therefore$ The standard deviation,

$$(\text{S.D.}) = \sqrt{Npq} \tag{4.7}$$

$E(X) = Np$ means that on the average we can expect Np successes. For example, if the probability of death from a particular disease is 0.01, and if 1000 people are affected by this disease, we can expect $Np = 1000 \times 0.01 = 10$ deaths on the average, in the sense that even though there may not be exactly 10 deaths in a given set of 1000 patients, but if we consider such batches of 1000 patients then on the average there may be 10 deaths per 1000 patients. The variance in this case is $Npq = 1000 \times 0{\cdot}01 \times 0{\cdot}99 = 9{\cdot}9$ and the standard deviation or a measure of dispersion is $(1000 \times 0{\cdot}01 \times 0{\cdot}99)^{1/2}$.

4.22. Moment Generating function.

$$M_x(t) = E e^{tX}$$

$$= \sum_{0 \leqslant x \leqslant N} e^{tx} \binom{N}{x} p^x \; q^{N-x}$$

$$= \sum_{0 \leqslant x \leqslant N} \binom{N}{x} (pe^t)^x \; q^{N-x}$$

$$= (q+pe^t)^N \left[\begin{array}{l} \text{Since } (q+pe^t)^N \text{ when expanded gives all the terms in} \\ \sum_{0 \leqslant x \leqslant N} \binom{N}{x} (pe^t)^x \; q^{N-x} \end{array} \right] \quad (4.8)$$

$\therefore$ For the Binomial distribution the M.G.F.,

$$M_x(t) = (q+pe^t)^N .$$

The various raw moments may be obtained by differentiation

$$\frac{d}{dt} M_x(t) = N(q+pe^t)^{N-1} \cdot pe^t.$$

$$\therefore \quad \frac{d}{dt} M_x(t) \Big|_{t=0} = N(q+p)^{N-1} p = Np \qquad (\text{Since } q+p=1) \quad (4.9)$$

$$\therefore \quad E(X) = \mu = \mu_1' = Np$$

$$\frac{d^2}{dt^2} M_x(t) \Big|_{t=0} = Np(q+p)^{N-1} + N(N-1)p^2(q+p)^{N-2}$$

$$= Np + N(N-1)p^2. \quad (4.10)$$

$\therefore$ The variance of $X = \sigma^2 = \mu_2' - \mu_1'^2$

$$= Np + N(N-1)p^2 - (Np)^2$$

$$= Np - Np^2$$

$$= Np(1-p) = Npq$$

$\therefore$ The standard deviation

$$\sigma = \sqrt{Npq}. \quad (4.11)$$

Higher order moments may be obtained by successive differentiation.

Ex. 4.22.1. *The moment generating function of a s.v. X is given as*

$$M_x(t) = \frac{1}{256} (3+e^t)^4,$$

find the probability function of X.

Sol.

$$M_x(t) = \frac{1}{256} (3+e^t)^4 = \frac{1}{4^4} (3+e^t)^4$$

$$= \left(\frac{3}{4} + \frac{1}{4} e^t \right)^4 .$$

This is of the form $(q+pe^t)^4$ where $q = 3/4$ and $p = 1/4$.

$\therefore$ By the uniqueness theorem, the corresponding distribution is Binomial with probability function $\binom{4}{x}(1/4)^x(3/4)^{4-x}$.

In the Binomial law, the stochastic variable X is the number of successes in a Binomial situation. If we consider the proportion of successes $\frac{X}{N}$ then the expected value and variance of the stochastic variable $Y=\frac{X}{N}$ may be easily obtained as,

$$E(Y)=E\left(\frac{X}{N}\right)=\frac{1}{N}E(X)$$

(Since N is a constant with respect to X)

$$=\frac{Np}{N}=p. \tag{4.12}$$

$$\text{Var }(Y)=\frac{1}{N^2}\text{ Var }(X)$$

[Since Var $(aX+b)=a^2$ Var (X)]

$$=\frac{1}{N^2}Npq=\frac{pq}{N} \tag{4.13}$$

$\therefore$ The standard deviation of

$$Y=(pq/N)^{1/2} \tag{4.14}$$

It is interesting to see that the factorial moments are easier to evaluate when X is a Binomial variate. For example the second factorial moment $\mu_{[2]}$, by definition, is

$$\mu_{[2]}=EX(X-1)$$

$$=\sum_{0\leqslant x\leqslant N} x(x-1)\binom{N}{x}p^x\ q^{N-x}$$

$$=N(N-1)p^2 \tag{4.15}$$

Similarly $\mu_{[r]}$ for $r=1, 2, 3,\ldots$ may be easily evaluated.

4.23. Recurrence formula. For convenience of evaluation of Binomial probabilities a recurrence formula may be obtained as follows.

$$f(x,\theta)=\frac{N!}{x!\,(N-x)!}p^x\ q^{N-x}$$

$$\therefore\quad f(x+1,\theta)=\frac{N}{(x+1)!\,(N-x-1)!}p^{x+1}\ q^{N-x-1}$$

(x is replaced by $x+1$)

$$\therefore \quad \frac{f(x+1,\theta)}{f(x,\theta)}=\frac{N-x}{x+1}\cdot\frac{p}{q}$$ (obtained by cancelling out all the common factors)

$$\therefore \quad f(x+1,\theta)=\frac{N-x}{x+1}\cdot\frac{p}{q}f(x,\theta) \tag{4.16}$$

From the formula if we know $f(0, \theta)$ and p we can evaluate $f(1, \theta), f(2, \theta), \ldots\ldots$

$$f(1,\theta)=\frac{N-0}{0+1}\,\frac{p}{q}\,f(0,\theta)$$

$$=N\frac{p}{q}\,f(0,\theta)$$

But $$f(0,\theta)=\binom{N}{0}p^0\,q^{N-0}=q^N=(1-p)^N.$$

Therefore, if $f(0, \theta)$ and p are given, N is obtained and thereby $f(1, \theta)$ is obtained and

$$f(2,\theta)=(N-1)\,f(1,\theta)/2.$$

From this $f(2, \theta)$ is obtained. Proceeding like this all other probabilities are obtained. For computational purposes the Binomial coefficients $\binom{N}{x}\;p^x\;\;q^{N-x}$ for various values of N, p and x are tabulated. Extracts of these tables are given at the end of this book. References to more tables are given at the end of this chapter.

Exercises

4.1. For a Binomial distribution with parameters $N=5$ and $p=0{\cdot}3$ find the probability of getting (*a*) atleast 3 successes, (*b*) at most 3 successes, (*c*) exactly 3 failures.

4.2. Find the probability of getting (*a*) exactly 4 heads, (*b*) atleast 3 tails, (*c*) atmost 3 heads, when an unbiased coin is thrown 6 times.

4.3. Assuming a Binomial probability situation with $p=1/2$ what is the probability that out of 20 babies born in a hospital 15 are boys ?

4.4. Show that for $p=1/2$ the Binomial distribution has a maximum at $x=N/2$ when N is even and at $x=(N-1)/2$ and $(N+1)/2$ when N is odd.

4.5. The probability that a bomb hits a target is given to be 0·80. Assuming a binomial situation, what is the probability that out of 10 bombings exactly 4 will be misses ?

4.6. In a game of taking a chace, a contestant has to give correct answers to 4 out of 5 questions to win the contest. Questions are given with 3 answers each, out of which one is a correct answer. If a contestant answers the questions by selecting the answers at random, what is the probability that he will win the contest ?

4.7. For the Binomial distribution with parameters p and N show that

$$\mu_{r+1}/pq=N.r.\mu_{r-1}+\frac{d}{dp}\mu_r,$$

where μ_r is the r^{th} central moment and $q=1-p$. Find μ_2 and μ_3 by using this recurrence relation.

4.8. A natural scientist claims that only 10% of the birds of a particular category are affected by radio-active fallout. A testing service tries to check this claim by choosing a random sample of 30 birds. The testing service will accept the claim if only 0 or 1 or 2 birds in the sample are affected, otherwise it will reject the claim. What are the probabilities that (1) the service will accept the claim when actually 15% of the birds are affected, (2) the service will reject the claim if the claim is true ?

4.9. A manufacturer of razor blades claims that only 4% of the blade do not meet the specified quality limits. A customer will accept purchases of 12 blades if 0 or 1 does not meet the quality limits ; otherwise he will return the purchase. Assuming that this is Binomial probability situation, and the purchase is a random sample of size 12, what are the probabilities that (1) he will return the purchase when the manufacturer's claim is true, (2) he will accept the purchase when actually 5% of the manufacturer's razor blades do not meet quality limits.

4.24. Poisson Distribution. The probability function for the Poisson distribution is seen to be (section 4.1),

$$f(x, \theta)=\lambda^x \; e^{-\lambda}/x\,!$$

for $$x=0, 1, \ldots\ldots, \infty$$

and $$f(x, \theta)=0,$$

elsewhere, where λ (lamda) is a parameter and

$$\lambda>0. \tag{4.17}$$

The Poisson probability law may be obtained as a limiting form of the Binomial probability function

$$f(x, \theta)=\binom{N}{x} p^x \; q^{N-x}.$$

Let us consider the siutation when $N\to\infty$, $p\to 0$ but Np remains a finite constant λ always ; *i.e.*, the number of trials is very very large, the probability of success in any trial is very very small but $Np=\lambda$ is a finite constant. This situation may be called a Poisson situation and we will show that in this situation the Binomial probability law tends to the Poisson probability law.

For the Binomial distribution

$$f(x, \theta)=\frac{N(N-1)\ldots(N-x+1)}{x\,!} \; p^x \; q^{N-x}$$

where $$q=1-p.$$

If $Np=\lambda$ then $P=\dfrac{\lambda}{N}$

$$\therefore \quad f(x, \theta)=\frac{1}{x\,!}\, N(N-1)\ldots(N-x+1) \left(\frac{\lambda}{N}\right)^x \left(1-\frac{\lambda}{N}\right)^{N-x}.$$

$$=\frac{1}{x!}\,\frac{N}{N}\,\frac{(N-1)}{N}\cdots\frac{(N-x+1)}{N}$$

$$\lambda^x\left(1-\frac{\lambda}{N}\right)^N\left(1-\frac{\lambda}{N}\right)^{-x}$$

$$=\frac{\lambda^x}{x!}1.\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{x-1}{N}\right)$$

$$\left(1-\frac{\lambda}{N}\right)^N\left(1-\frac{\lambda}{N}\right)^{-x} \quad (4.18)$$

when $N\to\infty$, $\left(1-\frac{1}{N}\right)\to 1$

$$\left(1-\frac{2}{N}\right)\to 1$$

$$\left(1-\frac{x-1}{N}\right)\to 1 \quad \text{(Since } x \text{ is finite)}$$

and $\left(1-\frac{\lambda}{N}\right)^N\to e^{-\lambda}$

$$\left[\text{Since } \lim_{n\to\infty}\left(1+\frac{t}{n}\right)^n=e^t\right]$$

and $\left(1-\frac{\lambda}{N}\right)^{-x}\to 1$

(Since x and λ are finite quantities)

$$\therefore \quad f(x,\theta)\to\frac{\lambda^x}{x!}e^{-\lambda} \quad (4.19)$$

where θ represents the only parameter λ.

By taking the limit of a Binomial probability function it is possible that the resulting limiting form may not satisfy the conditions for a probability function or the limiting form may not be a probability function. Let us examine whether $\frac{\lambda^x}{x!}e^{-\lambda}$ is a probability function or not. If $f(x)$ is to be a probability function, then the following conditions are to be satisfied.

(1) $f(x)\geqslant 0$

(2) $\int_{-\infty}^{\infty} f(x)dx=1$

or $\sum_{-\infty<x<\infty} f(x)=1$

Here $\frac{\lambda^x}{x!}e^{-\lambda}$ is evidently >0 Since $x=0, 1, 2, 3,\ldots$ and λ is a positive quantity ($\lambda=Np$).

$$\sum_{0\leqslant x<\infty} \frac{\lambda^x}{x\,!} e^{-\lambda} = e^{-\lambda} \sum_{0\leqslant x<\infty} \frac{\lambda^x}{x\,!}$$

$$= e^{-\lambda}\left[1+\frac{\lambda}{1\,!}+\frac{\lambda^2}{2\,!}+\ldots\right]$$

$$= e^{-\lambda}\; e^{\lambda} = 1$$

$$\left(\text{since } \; e^y = 1+\frac{y}{1\,!}+\frac{y^2}{2\,!}+\ldots\right)$$

$$\therefore \qquad \frac{\lambda^x}{x\,!} e^{-\lambda} = f(x) \text{ is a probability function.} \qquad (4.20)$$

This probability function is known as the Poisson probability function and we will denote this by our general notation

$$f(x, \theta) = \frac{\lambda^x}{x\,!} e^{-\lambda} \text{ for } x=0, 1, 2, \ldots$$

$$= 0 \text{ elsewhere.}$$

where $\lambda > 0$ is a parameter.

This was first introduced by a French mathematician called S. Poisson and hence it is called the Poisson distribution.

4.24.1. Moments of the Poisson Distributions.

$$E(X) = \sum_{0\leqslant x<\infty} x \frac{\lambda^x}{x\,!} e^{-\lambda} = \sum_{1\leqslant x<\infty} x \frac{\lambda^x}{x!} e^{-\lambda}$$

(when $x=0$ the corresponding term is zero)

$$= \lambda\, e^{-\lambda} \sum_{1\leqslant x<\infty} \frac{\lambda^{x-1}}{(y-1)\,!}$$

$$= \lambda\, e^{-\lambda}\left[1+\frac{\lambda}{1\,!}+\frac{\lambda^2}{2\,!}+\ldots\right]$$

$$= \lambda e^{-\lambda} \,.\, e^{\lambda} = \lambda \qquad (4.21)$$

The second factorial moment

$$\mu_{[2]} = EX(X-1) = E(X)^2 - E(X) = \mu'_2 - \mu$$

$$= \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x}{x!} e^{-\lambda}$$

$$= \lambda^2 \text{ (see Ex. 3.34.2)} \qquad (4.22)$$

$$\therefore \qquad \text{Var }(X) = \mu_2 = \mu'_2 - \mu^2$$

$$= [\mu_{[2]} + \mu] - \mu^2$$

$$= \lambda^2 + \lambda - \lambda^2 = \lambda \qquad (4.23)$$

$\therefore$ The standard deviation $= \sqrt{\lambda}$.

$\therefore$ For a Poisson distribution the mean $=$ the variance $=\lambda$.

The reader may evaluate μ_3 and μ_4.

4.24.2. The moment generating function for the Poisson distribution.

$$M_x(t)=Ee^{tX}$$

$$=\sum_{x=0}^{\infty} e^{tx}\frac{\lambda^x}{x\,!}\,e^{-\lambda}$$

$$=e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x\,!}$$

$$=e^{-\lambda}\left[1+\frac{\lambda e^t}{1\,!}+\frac{(\lambda e^t)^2}{2\,!}+\cdots\right]$$

$$=e^{-\lambda}.\ e^{\lambda e^t}$$

$$=e^{\lambda\,(e^t-1)} \tag{4.24}$$

By differentiating

$$M_x(t)=e^{\lambda(e^t-1)}$$

with respect to t the various moments can be easily obtained.

The probability of getting exactly x successes in a Poisson situation is

$$f(x,\theta)=\frac{\lambda^x}{x\,!}\quad{}^{-\lambda}\ .$$

A recurrence relation may be obtained since

$$f(x+1,\theta)=\frac{\lambda}{x+1}\,f(x,\theta).$$

We have seen that a binomial distribution approximates to a Poisson distribution when $N\to\infty$, $p\to 0$ but $Np=\lambda$, a constant. If $p\to 0$ the probability of a success in any trial is very very small. So the Poisson probability law is sometimes called the probability law for rare events. This law may be applied to situations like death due to snake bite, delivery of quintuplets, emission of alpha particles, traffic accidents etc. Since when $N\to\infty$, $p\to 0$ and $Np=\lambda$, is a constant, the Binomial law becomes a Poisson law, in practical situations we evaluate the probabilities by the Binomial law only when N is not sufficiently large and p or q does not tend to zero.

Ex. 4.24.1. *A hospital switch board receives an average of 4 emergency calls in a 10 minute interval. What is the probability that (1) there are at the most 2 emergency calls in a 10 minute interval (2) there are exactly 3 emergency calls in a 10 minute interval ?*

Sol. The distribution of the number of emergency calls may be considered to be a Poisson distribution with mean

$$=E(X)=\lambda=4$$

in a unit time of 10 minutes.

$\therefore$ The probability of getting exactly x emergency calls in a unit time of 10 minutes

$$=f(x,\theta)=\frac{\lambda^x}{x!}e^{-\lambda}=\frac{4^x}{x!}e^{-4}.$$

(1) The probability of getting at the most 2 emergency calls

=Prob. of getting 0 or 1 or 2 calls

$$=\frac{4^0}{0!}e^{-4}+\frac{4^1}{1!}e^{-4}+\frac{4^2}{2!}e^{-4}$$

$$=e^{-4}[1+4+8]$$

$$=13 \,.\, e^{-4}=0\cdot2579.$$

(2) The probability of getting exactly 3 emergency calls

$$=\frac{4^3}{3!}e^{-4}=\frac{32}{3}e^{-4}=0\cdot1952.$$

Ex. 4.24.2. *Suppose that the probability of death in case of influenza is 0.01. If 20 influenza patients are there in a hopital what is the probability that exactly 2 patients will die ? Approximate the probability by a Poisson distribution and evaluate the error in this approximation.*

Sol. This may be taken as a Binomial situation with

$$N=20 \quad p=0.01$$

and
$$q=0\cdot99$$

$\therefore$ The probability that exactly 2 patients will die is

$$=\binom{20}{2}(0.01)^2(0.99)^{18}$$

$$=0.0158$$

If the Binomial probabilities are approximated by Poisson probabilities

then
$$\lambda=Np=20\times0.01=0\cdot2$$

$\therefore$ The probability of getting exactly 2 successes

$$=\frac{\lambda^2}{2!}e^{-\lambda}$$

$$=\frac{(0.2)^2}{2\,!}e^{-0.2}=0.0164$$

$\therefore$ The error in the approximation

$$=0.0164-0.0158=0.0006$$

Comments. For practical purposes a good Poisson approximation to the Binomial probabilities is obtained for a N which is as small as 20 provided $Np<5$. Poisson probabilities are easier to evaluate. So when N is sufficiently large and p is sufficiently small a Poisson approximation may be enough for practical purposes. Poisson probabilities are tabulated for various values of λ and x. A table is given at the end of this book and references are given at the end of this chapter.

A Poisson distribution may also be obtained independently (*i.e.*, without considering it as a limiting form of the Binomial distribution). Consider the following experimental situation :

(1) The probability of a success in a small time interval from t to $t+\Delta t$ is $\alpha . \Delta t$ where α is a positive constant and Δt denotes a small increment in time at t.

(2) The probability of getting more than one success in this interval is negligibly small. (we will assume this to be zero).

(3) The probability of a success in interval t to $t+\Delta t$ does not depend on the success or failure prior to time t. Under these three conditions it may be shown that the probability of getting exactly x successes in the time t, say, $f(x, t)$ is given by

$$f(x, t)=\frac{\lambda^x}{x!}\ e^{-\lambda} \text{ for } x=0, 1, 2,... \tag{4.25}$$

where $\lambda=\alpha . t$ and α and t are both positive quantities.

The event of getting exactly x successes in time $t+\Delta t$ may be partitioned into two mutually exclusive events of (1) getting exactly x successes in time t and then a failure in the interval t, $t+\Delta t$, (2) getting exactly $x-1$ successes in time t and a success in this interval t to $t+\Delta t$.

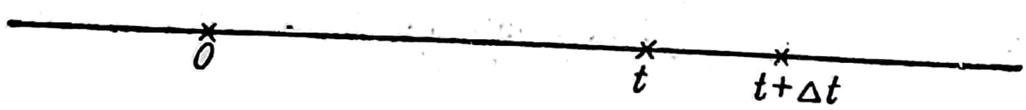


Fig. 4.1.

$f(x, t)$=probability of getting exactly x successes in time t.

$f(x, t+\Delta t)$=probability of getting exactly x successes in time $t+t\Delta$.

$f(x-1, t)$=probability of getting exactly $x-1$ successes in time t.

Probability of a success in time t to $t+\Delta t=\alpha . \Delta t$.

$\therefore$ The probability of a failure in this interval

$$=1-\alpha . \Delta t.$$

(We assumed that the prob. of getting more than one success is negligible).

$$\therefore \qquad f(x, t+\Delta t)=f(x-1, t)\,\alpha\,.\,\Delta t+f(x, t)(1-\alpha\,.\,\Delta t)$$
$$=f(x, t)+f(x-1, t)\alpha.\,\Delta t-f(x, t)\,\alpha.\,\Delta t \tag{4.26}$$

$$\frac{f(x, t+\Delta t)-f(x, t)}{\Delta t}=\alpha[f(x-1, t)-f(x, t)] \tag{4.27}$$

Taking limits when $\Delta t \to 0$

$$\frac{d}{dt} f(x, t)=\alpha[f(x-1, t)-f(x, t)] \tag{4.28}$$

Solving this differenco-differential equation it can be shown

that $$f(x, t)=\frac{\lambda^x}{x\,!}\,e^{-\lambda}$$

where $$\lambda=\alpha t,$$
$$x=0, 1, 2, \ldots$$

The proof is beyond the scope of this book.

Ex. 4.24.3. *Alpha particles are emitted by a radioactive source at the rate of 2 per every 5 minutes on the average. Assuming this as a Poisson situation, what is the probability of getting exactly 4 emissions in 15 minutes.*

Sol. Here 5 minutes is a unit of time. So we have to find out the probability of getting 4 successes in 3 units of time. According to the notation, $\alpha=2$ and $t=3$ and $x=4$.

$\therefore$ The required probability

$$=\frac{(\lambda\alpha t)^x}{x\,!}\,e^{-\alpha t}=\frac{6^4}{4!}\,e^{-6}=0.1339.$$

Exercises

4.10. For a Binomial distribution with parameters $N=50$ and $p=0.1$ compare the Binomial probabilities to that of the approximating Poisson probabilities for $x=0, 1, 2, 3, 4$ and 5.

4.11. Show that the M.G.F. of a Binomial distribution approaches the M.G.F. of a Poisson distribution when $N\to\infty$, $p\to 0$ but $Np=\lambda$ is a constant. Hence show that the Binomial distribution tends to a Poisson distribution under these conditions

4.12. A machine producing electric bulbs is known to produce 1% defectives. What is the probability of getting 3 defective bulbs if a random sample of size 20 is chosen ? Approximate this probability by a Poisson distribution.

4.13. The following moment generating functions are given. Use the uniqueness theorem to determine the corresponding probability distributions.

(*a*) $(1/2+e^t/2)^5$, (*b*) $(2+e^t)^3/27$, (*c*) exp. $2(e^t-1)$.

4.14. Differentiating $\mu_r = E(X-\lambda)^r$ for a Poisson distribution with parameter λ, obtain the recurrence formula,

$$\mu_{r+1}/\lambda = r\mu_{r-1} + \frac{d}{d\lambda}\mu_r.$$

Find μ_2 and μ_3.

4.15. The number of traffic accidents in a city each month is assumed to be a Poisson variate with the parameter $\lambda=4$. What is the probability that (*a*) there are 6 accidents in a certain month, (*b*) there are fewer than 4 accidents in a certain month.

4.16. An office switch board receives telephone calls at the rate of 3 calls per minute on the average. What is the probability of receiving (*a*) no calls in a one minute interval, (*b*) at the most 3 calls in a 5 minute interval.

4.17. The probability that a vehicle will pass through a particular point in a small interval of time Δt is $0.1\,\Delta t$, time being measured in minutes. What is the probability that a traffic counting device at that point will show (*a*) exactly 5 counts in a one minute interval, (*b*) at the most 5 counts in a 20 minute interval.

4.3. THE HYPERGEOMETRIC DISTRIBUTION

If the binomial probability law is to be applied to an experimental situation then the situation should be a binomial probability situation. If the probability of success in any trial does not remain the same from trial to trial then the binomial law can not be applied. In this case we will see that a hypergeometric distribution is appropriate.

Let us consider the following situation. Suppose that we have $a+b$ objects out of which a are of one type and b are of a second type. n objects are taken at random from this set of $a+b$ objects. The probability that x of the n objects taken, belong to the set containing a objects and the remaining $n-x$ objects belong to the second set containing b objects, is evidently

$$\frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}.$$

Thus
$$f(x,\theta) = \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}$$

for
$$x=0,1,2,\ldots n \text{ or } a \text{ and } n-x \leqslant b. \qquad (4.29)$$
$$=0 \text{ elsewhere.}$$

is called the hypergeometric distribution where a, b, n are parameters ; *i.e.*, for different values of a, b, n different hypergeometric distributions are obtained, and for specific values of a, b and n (for example $a=10$, $b=20$, $n=13$) the hypergeometric distribution is completely specified. It can be easily verified that

$$\sum_{x=0}^{n} \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}} = 1.$$

Hypergeometric distribution may also be considered to be the probability law when sampling without replacement. If objects are taken one by one without replacement then the probability of getting exactly x objects belonging to one type consisting of a objects and $n-x$ objects belonging to a second type consisting of b objects, is evidently given by the hypergeometric probability law, if there are only $a+b$ objects and n are taken without replacement.

Ex. 4.3.1. *In a basket of 100 tomatoes 20 are rotten. Fifteen tomatoes are picked up at random one by one. (Here at random does not mean haphazardly, but it means that when one tomato is taken every tomato in the basket is given equal chances of being taken). What is the probability that*

(1) *there are exactly 2 rotten tomatoes:*

(2) *there are at the most 2 rotten tomatoes,*

(3) *there are at least 2 rotten tomatoes in the sample ?*

Sol. The 100 tomatoes may be classified into two types (80 good and 20 bad).

(1) The required probability

$$= \frac{\binom{20}{2}\binom{80}{13}}{\binom{100}{15}}.$$

(2) Atmost 2 rotten tomatoes means 0 or 1 or 2 bad ones.

$\therefore$ The required probability

$$= \frac{\binom{20}{0}\binom{80}{15}}{\binom{100}{15}} + \frac{\binom{20}{1}\binom{80}{14}}{\binom{100}{15}} + \frac{\binom{20}{2}\binom{80}{13}}{\binom{100}{15}}$$

(3) At least 2 rotten tomatoes means 2 or 3 or... or 15 bad ones

$\therefore$ The required probability

$$= \frac{\binom{20}{2}\binom{80}{13}}{\binom{100}{15}} + \frac{\binom{20}{3}\binom{80}{12}}{\binom{100}{15}} + \dots + \frac{\binom{20}{15}\binom{80}{0}}{\binom{100}{15}}$$

=the probability of not getting exactly 0 or 1 rotten ones

=1 − prob. of getting 0 or 1 bad ones

$$=1-\frac{\binom{20}{0}\binom{80}{15}}{\binom{100}{15}}-\frac{\binom{20}{1}\binom{80}{14}}{\binom{100}{15}}.$$

Comments. These different probabilities may be completely evaluated by the help of a table of Binomial coefficients or by using logarithms. These computations are left to the reader.

Ex. 4.3.2. *From a well shuffled deck of 52 playing cards 2 cards are selected at random without replacement. What is the probability of getting 2 aces ?*

Sol. There are 4 aces altogether. So these 52 cards may be divided into two types, namely a set of 4 aces and a set of 48 non-aces.

∴ The required probability

$$=\frac{\binom{4}{2}\binom{48}{0}}{\binom{52}{2}}=\frac{4\times 3}{52\times 51}=\frac{1}{221}$$

4.31. Moments.

$$\mu=E(X)=\sum_{x=0}^{n} x\frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}$$

$$=\frac{1}{\binom{a+b}{n}}\sum_{x=0}^{n} x\frac{a\,!}{x\,!\,(a-x)}\binom{b}{n-x}$$

$$=\frac{1}{\binom{a+b}{n}}\sum_{x=0}^{n} x\frac{a\,!}{x\,!\,(a-x)\,!}\binom{b}{n-x}$$

(when $x=0$ the corresponding term is zero)

$$=\frac{1}{\binom{a+b}{n}}\sum_{x=1}^{n} \frac{a\,!}{(x-1)\,!\,(a-x)\,!}\binom{b}{n-x}$$

Put $a-1=A$, $x-1=y$ and $n-1=N$, then E(X) may be written as,

$$E(X)=\frac{a}{\binom{a+b}{n}}\sum_{y=0}^{N}\frac{A\,!}{y\,!\,(A-y)\,!}\binom{b}{N-y}$$

$$=\frac{a}{\binom{a+b}{n}}\sum_{y=0}^{N}\binom{A}{y}\binom{b}{N-y} \tag{4.30}$$

$$=\frac{a}{\binom{a+b}{n}}\binom{A+b}{N}$$

$\left(\text{Since } \sum_{t=0}^{r}\binom{a}{t}\binom{b}{r-t}=\binom{a+b}{r}\right.$

This may be seen by comparing the coefficients of x^r on both sides of

$\left.(1+x)^a(1+x)^b=(1+x)^{a+b}\right)$

$$=\frac{a}{\binom{a+b}{n}}\binom{a+b-1}{n-1}=n\cdot\frac{a}{(a+b)}$$

$$\therefore \qquad E(X)=n\frac{a}{a+b} \tag{4.31}$$

when X is a hypergeometric variate.

The variance of X may be easily obtained from the second factorial moment $\mu_{[2]}=EX(X-1)$.

$$\mu_{[2]}=EX(X-1)=\sum_{x=0}^{n}x(x-1)\frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}$$

$$=\sum_{x=2}^{n}x(x-1)\frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}$$

$$=\sum_{x=2}^{n}x(x-1)\frac{a\,!}{x\,!\,(a-x)\,!}\frac{\binom{b}{n-x}}{\binom{a+b}{n}}$$

$$=a(a-1)\sum_{x=2}^{n}\frac{(a-2)\,!}{(x-2)\,!\,(a-x)\,!}\frac{\binom{b}{n-x}}{\binom{a+b}{n}} \tag{4.32}$$

Put $a-2=\text{A}$, $x-2=y$, $n-2=\text{N}$ then $\mu_{[2]}$ becomes

$$\mu_{[2]}=\frac{a(a-1)}{\binom{a+b}{n}}\sum_{y=0}^{\text{N}}\frac{\text{A}\,!}{y\,!\,(\text{A}-y)}\binom{b}{\text{N}-y}$$

$$=\frac{a(a-1)}{\binom{a+b}{n}}\sum_{y=0}^{\text{N}}\binom{\text{A}}{y}\binom{b}{\text{N}-y}$$

$$=\frac{a(a-1)}{\binom{a+b}{n}}\binom{\text{A}+b}{\text{N}}$$

$$=\frac{a(a-1)}{\binom{a+b}{n}}\binom{a+b-2}{n-2}$$

$$=\frac{a(a-1)n(n-1)}{(a+b)(a+b-1)}\,. \tag{4.33}$$

But

$$\mu_{[2]}=\text{EX}(\text{X}-1)=\text{E}(\text{X})^2-\text{E}(\text{X})$$

$\therefore$

$$\text{E}(\text{X}^2)=\mu_{[2]}+\text{E}(\text{X})$$

$$=\frac{a(a-1)n(n-1)}{(a+b)(a+b-1)}+\frac{a\,.\,n}{(a+b)}$$

But $\quad \text{Var (X)}=\mu_2=\text{E}(\text{X}^2)-(\text{EX})^2$

$$=\frac{a(a-1)\,n(n-1)}{(a+b)(a+b-1)}+\frac{a\;n}{(a+b)}-\left(\frac{a\;n}{a+b}\right)^2$$

$$=\frac{n\,ab\,.\,(a+b-n)}{(a+b)^2(a+b-1)} \tag{4.34}$$

$\therefore$ The standard deviation is

$$\sqrt{\mu_2}=\left[\frac{n\,ab(a+b-n)}{(a+b)^2(a+b-1)}\right]^{\frac{1}{2}} \tag{4.35}$$

In Ex. 4.3.1, we can expect on the average

$$\frac{a\,.\,n}{a+b}=\frac{20\times 15}{20+80}=3$$

bad tomatoes, in a random sample of 15 tomatoes from the basket. In the Binomial situation E(X) is seen to be $\text{E}(\text{X})=\text{N}p$. In

a Hypergeometric situation $E(X)$ is $\frac{a}{a+b}n$. It may be easily verified that a Hypergeometric distribution may be approximated to a Binomial distribution by taking $p=\frac{a}{a+b}$. In Ex. 4.3.1 if we use this approximation the probability of getting exactly 2 bad tomatoes from the lot is approximately equal to

$$\binom{N}{x} p^x q^{N-x} = \binom{15}{2}\left(\frac{20}{20+80}\right)^2\left(\frac{80}{100}\right)^{13}$$

$$=\binom{15}{2}\left(\frac{20}{100}\right)^2\left(\frac{80}{100}\right)^{13}$$

$$=0.231$$

The exact probability is

$$\frac{\binom{0}{2}\binom{80}{13}}{\binom{100}{15}}=0.152.$$

A recurrence relation may be obtained for a Hypergeometric distribution also.

4.4. THE NEGATIVE BINOMIAL DISTRIBUTION

Here we will consider the Binomial probability situation with a slight modification. Consider the situation where (1) the trials are independent, (2) the probability of success p in a trial remains the same from trial to trial. Suppose that we are interested in finding out the probability of getting the k^{th} success at the x^{th} trial. Here, evidently, the number of trials is not a constant. The required probability may be obtained by considering the events of getting exactly $k-1$ successes in $x-1$ trials, and subsequently the x^{th} trial results in a success ; *i.e.*, the required probability

$$=\binom{x-1}{k-1} p^{k-1} q^{(x-1)-(k-1)} p$$

$$=\binom{x-1}{k-1} p^k q^{x-k} \text{ for } x=k, k+1,\ldots$$

$$\therefore \quad f(x,\theta)=\binom{x-1}{k-1} p^k q^{x-k} \text{ for } x=k, k+1,\ldots \tag{4.36}$$

$$=0 \text{ elsewhere,} \qquad 0<p<1, \; q=1-p.$$

gives the probability law, called the Negative Binomial Distribution. Here p and k are parameters. The various probabilities may be easily seen to be the different terms in the binomial expansion of

$$\left(\frac{1}{p}-\frac{q}{p}\right)^{-k}=p^k(1-q)^{-k}$$

Hence the distribution is called a Negative Binomial Distribution.

Ex. 4.4.1. *If a boy is throwing stones at a target, what is the probability that his 10th throw is the 5th hit, if the probability of hitting the target at any trial is 0.40?*

Sol. This is a Negative Binomial Situation. According to the above notation $x=10$, $k=5$, and $p=0.40$.

$\therefore$ The required probability

$$=\binom{x-1}{k-1}p^k q^{x-k}$$

$$=\binom{9}{4}(0\cdot40)^5(0\cdot60)^5=0\cdot1004$$

4.41. Geometric Distribution. In the negative binomial distribution, if $k=1$, *i.e.*, the probability of the number of trials required to get the first success or the probability that the x^{th} trial is the first success is given by

$$f(x,\ \theta)=pq^{x-1} \text{ for } x=1,\ 2,\ 3,\ldots \tag{4.37}$$

$$=0 \text{ elsewhere, } 0<p<1,\ q=1-p.$$

This probability distribution is called the geometric distribution. This may be derived independently by considering a Binomial situation where the number of trials is not fixed. Then the probability that the x^{th} trial results in the first success is given by the geometric distribution. The various probabilities for $x=1, 2,\ldots$ are the various terms of a geometric progression, and hence, the distribution is called a geometric distribution.

Ex. 4.41.1. *In Ex. 4.4.1 what is the probability that the 4th attempt is the first hit?*

Sol. Here $x=4$, $p=0.40$

and therefore the required probability

$$=0.40\ (0\cdot60)^3$$

$$=0\cdot0864.$$

Negative binomial and geometric probabilities may be evaluated by using tables for factorials, binomial probabilities, logarithms etc. If tables are not available recurrence formula will

help in evaluating various probabilities. Some of the recurrence relationships are given in the following table.

Name	The Probability function	The Recurrence Relationship
1. The Binomial distribution	$f(x,\theta)=\binom{N}{x}p^x q^{N-x}$ for $x=0, 1, 2, \ldots n$; $=0$ elsewhere	$f(x+1, \theta)=\frac{N-x}{x+1}\,\frac{p}{q}\,f(x, \theta)$
2. The Poisson distribution	$f(x, \theta)=\frac{\lambda^x}{x\,!}\,e^{-\lambda}$ for $x=0, 1, 2\ldots$; $=0$ elsewhere	$f(x+1, \theta)=\frac{\lambda}{x+1}\,f(x, \theta)$
3. The Hyper-geometric distribution	$f(x, \theta)=\frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}$ for $x=0, 1, 2, \ldots n$; $=0$ elsewhere	$f(x+1, \theta)=\frac{(n-x)(a-x)}{(x+1)(b-n+x+1)}\,f(x, \theta)$
4. The Negative Binomial distribution	$f(x, \theta)=\binom{x-1}{k-1}p^k q^{x-k}$ for $x=k, k+1\ldots$; $=0$ elsewhere	$f(x+1, \theta)=\frac{x\,q}{x-k+1}\,f(x, \theta)$
5. The Geometric distribution	$f(x, \theta)=pq^{x-1}$ for $x=1, 2, 3,\ldots$; $=0$ elsewhere	$f(x+1, \theta)=q\,f(x, \theta)$
6. The Discrete uniform distribution	$f(x, \theta)=\frac{1}{n}$ for $x=x_1, x_2, \ldots x_n$; $=0$ elsewhere	$f(x+1, \theta)=f(x, \theta)$

Exercises

4.18. Obtain the moment generating function for (*a*) a discrete uniform distribution with parameter n, (*b*) a geometric distribution with parameter p and identify the distributions with the following M.G.F's :

(1) $e^t(5-4e^t)^{-1}$, (2) $(1/2)e^t/(1-e^t/2)$.

4.19. If in a small township of 1000 people 40% are conservatives and 60% are liberals, what is the probability that in a random sample of 100 people from this township 90% are conservatives and 10% are liberals ?

4.20. The probability that a swimmer will succeed in swimming across a lake is 0.4. What is the probability that the 10th swimmer is (*a*) the first one to cross the lake, (*b*) the 4th one to cross the lake ?

4.21. The probability that a prediction of a soothsayer will come true is 0.01. What is the probability that his 20*th* prediction is the 4*th* one to come true ?

4.22. The logarithmic distribution is given by

$$f(x, \theta)=\frac{-(1-p)^x}{x \log p} \quad \text{for } x=1, 2, 3,...$$
$$=0 \text{ elsewhere and } 0<p<1.$$

Find E(X) for this distribution.

4.23. (Random Walk) A point moves along a straight line in jumps of one unit each, starting from a given point 0. The point takes a jump to the right or to the left with probabilities p and $1-p$ respectively. Each jump is independent of all other jumps. If x is the distance from 0 after N jumps, show that

$$f(x)=\binom{N}{\frac{x+N}{2}} (pq)^{N/2}\left(\frac{p}{q}\right)^{x/2}$$

for $x= .. -2, -1, 0, 1, 2,......$ where $q=1-p$.

Assume $\binom{N}{r}=0$ if r is not an integer. (For more problems of this nature the reader may refer to, W. Feller, An Introduction to Probability Theory and Its Applications, John Wiley and sons, New York, 1957).

4.5. RECTANGULAR OR UNIFORM DISTRIBUTION

This is a simple continuous probability distribution with density function

$$f(x, \theta)=\frac{1}{\beta-\alpha} \text{ for } \alpha<x<\beta, \ \alpha>0 \text{ and } \alpha<\beta, \quad (4.38)$$
$$=0 \text{ elsewhere.}$$

In this distribution $\theta=(\alpha, \beta)$, or there are two parameters α and β. Fig. 4.2 gives a graphical representation of the distribution

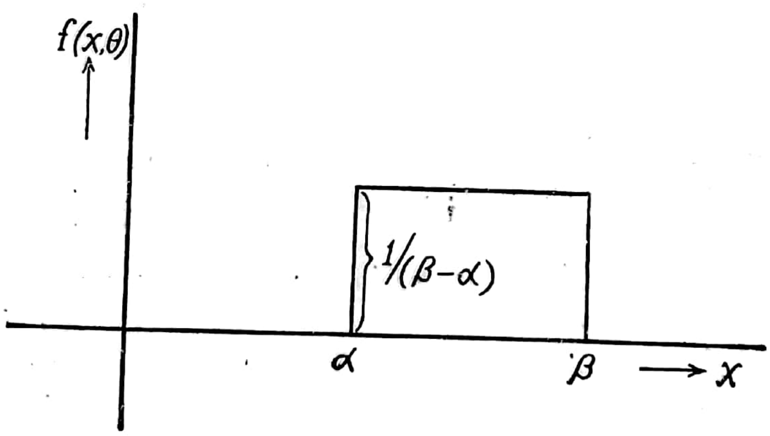


Fig. 4.2

Because of the rectangular shape of the distribution it is called a rectangular distribution.

4.5.1. Moments.

$$\mu'_r=E(X^r)=\int_{-\infty}^{\infty} x^r f(x, \theta)dx$$

$$=\int_{\alpha}^{\beta} x^r \frac{dx}{\beta-\alpha}=\frac{\beta^{r+1}-\alpha^{r+1}}{(\beta-\alpha)(r+1)} \tag{4.39}$$

For $r=1, 2,...$ various raw moments are obtained.

When $\alpha=0$, a rectangular distribution with one parameter is obtained. For a rectangular distributions the distribution function $F(x)$ is given by,

$$F(x)=\int_{-\infty}^{x} f(x)dx$$

$$=0+\int_{\alpha}^{x} \frac{1}{\beta-\alpha}\,dx=\frac{x-\alpha}{\beta-\alpha}. \tag{4.40}$$

If X is distributed according to the rectangular distribution given in Section 4.5 then, the probability that X is greater than or equal to d where d is a given constant, is given by

$$P\{x \geqslant d\}=\int_{d}^{\infty} f(x)dx$$

$$=\int_{d}^{\beta} \frac{1}{\beta-\alpha}\,dx+0=\frac{\beta-d}{\beta-\alpha} \tag{4.41}$$

Such a probability statement has great significance in testing statistical hypotheses, which will be discussed later.

Ex. 4·5·1. *The dial of a spinner is marked 1 to 100. The spinner is completely balanced in the sense that the indicator, when rotated, is as likely to stop at a number as at any other number. What is the probability that in 2 out of 3 trials the indicator stops in between 20 and 30 ?*

Sol. If x denotes the distance of the stopping point from zero (see also Ex. 3·11·3) then X has a rectangular distribution,

$f(x)=1/100$ for $0<x<100$ and $f(x)=0$ elsewhere.

The probability that the indicator stops between 20 and 30 at any trial is

$$=\int_{20}^{30} f(x)\,dx=\int_{20}^{30} (1/100)\,dx=(1/10)=p(\text{say}).$$

Since p remains the same for any trial, the required probability is given by a Binomial probability law and is

$$=\binom{3}{2}p^2(1-p)^1=3(1/10)^2\,(9/10)=27/1000.$$

4·6. THE EXPONENTIAL DISTRIBUTION

The density function of an exponential distribution is given by,

$$f(x,\,\theta)=(1/\theta)\,e^{-x/\theta} \qquad \text{for } x>0,\ \theta>0$$
$$=0 \text{ elsewhere.}$$

Here θ is the only parameter. A graphical representation of the distribution is given in Fig. 4·3.

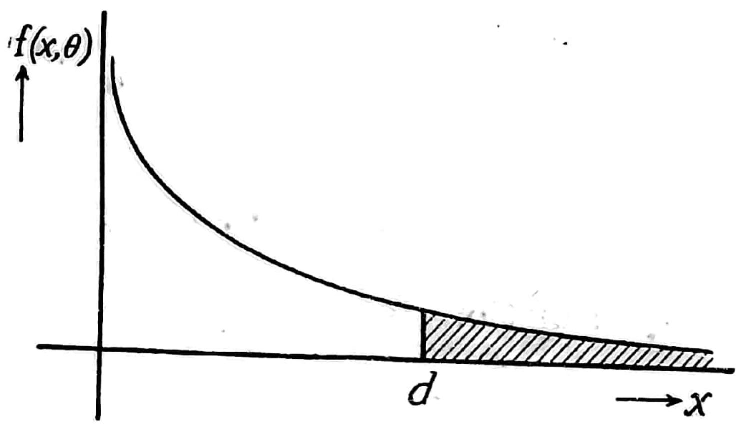


Fig. 4·3

The distribution function F(x) is given by,

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{0}^{x}(1/\theta)\,e^{-x/\theta}\,dx=1-e^{-x/\theta} \tag{4·43}$$

$$\text{Therefore, } 1-F(x)=\int_{x}^{\infty} f(x)\,dx=e^{-x/\theta}. \tag{4·44}$$

$$P\{x\geqslant d\}=\int_{d}^{\infty} f(x)\,dx=e^{-d/\theta}. \tag{4·45}$$

This is given by the shaded area in Fig. 4·3.

Ex. 4·6·1. *The daily consumption of milk in a city, in excess of 10,000 gallons, is approximately exponentially distributed with $\theta=1000$. The city has a daily stock of 20,000 gallons. What is the*

probability that the stock is insufficient for both the days if two days are selected at random ?

Sol. Let y denote the consumption on any day

$$X = Y - 10{,}000$$

has an exponential distribution,

$$f(x) = (1/1000)\ e^{-x/1000} \text{ for } 0 < x < \infty \text{ and } 0 \text{ elsewhere.}$$

The stock is insufficient if the demand exceeds the stock, that is,

When $x > 20{,}000 - 10{,}000 = 10{,}000$.

The probability that the stock is insufficient on any particular day,

$$= P\{x > 10{,}000\} = \int_{10{,}000}^{\infty} (1/1000)\ e^{-x/1000}\, dx = e^{-10}.$$

The required probability $= (e^{-10})^2 = e^{-20}$ (follows from independence).

4·7. THE GAMMA DISTRIBUTION

The density function for this probability distribution is given as,

$$f(x, \theta) = k\, x^{\alpha-1}\, e^{-x/\beta} \qquad \text{for } x > 0,\ \alpha,\ \beta > 0 \qquad (4{\cdot}46)$$
$$= 0 \text{ elsewhere.}$$

k is such that $f(x, \theta)$ is a density function. k may be evaluated by using the result that for any density function $f(x, \theta)$,

$$\int_{-\infty}^{\infty} f(x, \theta)\, dx = 1$$

i.e.,
$$\int_{-\infty}^{\infty} f(x, \theta)\, dx = \int_{0}^{\infty} k.\ x^{\alpha-1}\, e^{-x/\beta}\, dx$$

Make the substitution $t = x/\beta$ then $dt = dx/\beta$

$$\therefore \int_{0}^{\infty} x^{\alpha-1}\ e^{-x/\beta} dx = k.\ \beta^{\alpha} \int_{0}^{\infty} t^{\alpha-1}\, e^{-t}\, dt$$

$$= k\, \beta^{\alpha}\, \Gamma(\alpha)$$

$$\therefore \qquad k = \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)} \tag{4·47}$$

where $\Gamma(\alpha)$ is the Gamma function.

The Gamma distribution is an important distribution. When $\alpha=1$ the exponential distribution is obtained. There are practical situations where this distribution is applicable. Some of the results are given in the exercises at the end of this section. When $\alpha=n/2$ and $\beta=2$ a Chi-square distribution is obtained. A graphical representation of the Gamma distribution for various values of the parameters α and β, is given in Fig. 4·4.

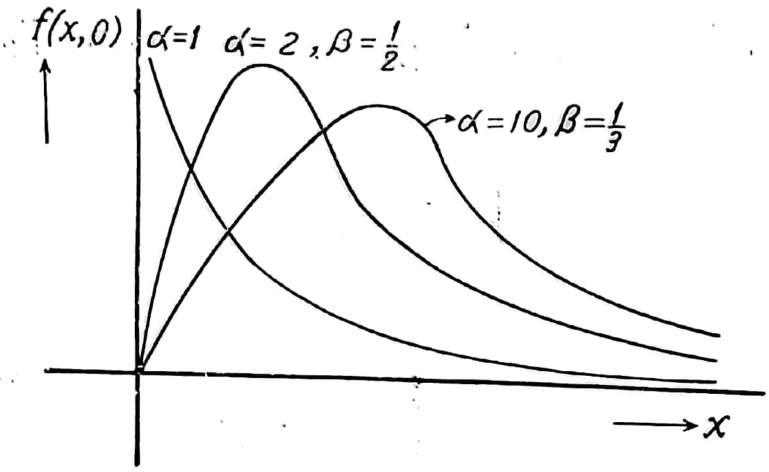


Fig. 4·4.

4·71. Moments. The various raw moments may be easily evaluated

$$\mu'_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x, \theta)\, dx$$

$$= \int_0^{\infty} \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)} x^r . x^{\alpha-1} e^{-x/\beta}\, dx$$

$$= \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)} \int_0^{\infty} x^{r+\alpha-1} e^{-x/\beta}\, dx.$$

Let us make a substitution $t=x/\beta$, then $dt=dx/\beta$

and
$$\mu'_r = \frac{\beta^{\alpha+r}}{\beta^{\alpha}\,\Gamma(\alpha)} \int_0^{\infty} t^{r+\alpha-1} e^{-t}\, dt$$

$$= \frac{\beta^r}{\Gamma(\alpha)} \Gamma(\alpha+r) \tag{4·48}$$

For various values of r the various raw moments are obtained ;

$$\text{when } r=1,\ \mu'_1=\frac{\beta}{\Gamma(\alpha)}\Gamma(\alpha+1)=\frac{\beta.\alpha.\Gamma(\alpha)}{\Gamma(\alpha)}=\alpha\beta. \tag{4·49}$$

$$\text{when } r=2,\ \mu'_2=\frac{\beta^2}{\Gamma(\alpha)}\Gamma(\alpha+2)=\frac{\beta^2(\alpha+1)\alpha.\Gamma(\alpha)}{\Gamma(\alpha)}$$

$$=\beta^2\alpha(\alpha+1). \tag{4·50}$$

$$\therefore \quad \text{Var}(X)=\mu_2=\mu'_2-\mu'^2_1$$

$$=\beta^2\alpha(\alpha+1)-\alpha^2\beta^2=\alpha\beta^2 \tag{4.51}$$

The distribution function $F(x)$ for a Gamma distribution is given by

$$F(x)=\int_{-\infty}^{x} f(x,\theta)dx$$

$$=1-\int_{x}^{\infty}\frac{1}{\beta^\alpha\,\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}dx$$

$$=1-\int_{t\beta}^{\infty}\frac{t^{\alpha-1}}{\Gamma(\alpha)}e^{-t}\,dt \tag{4.52}$$

The integral is very difficult to evaluate for a general α.

The integral $\displaystyle\int_{u}^{\infty}\frac{1}{\Gamma(\alpha)}t^{\alpha-1}e^{-t}\,dt$ is tabulated for various values of u and α. These tables are called the incomplete Gamma tables. A reference is given at the end of this chapter. By using an incomplete Gamma table we can evaluate the tail areas

$$\left(i.e.\ \int_{u}^{\infty}\frac{1}{\Gamma(\alpha)}t^{\alpha-1}e^{-t}\,dt\right)$$

or the distribution function

$$F(u)=\int_{0}^{u}\frac{1}{\Gamma(\alpha)}t^{\alpha-1}e^{-t}\,dt$$

of a Gamma distribution.

4.72. The Moment Generating Function.

$$M_X(t)=Ee^{tX}=\int_0^\infty c^{tx}\frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}\,dx$$

$$=\frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty x^{\alpha-1}\;e^{-x}\left(\frac{1}{\beta}-t\right)dx$$

Let $u=x\left(\frac{1}{\beta}-t\right)$ then $du=dx\left(\frac{1}{\beta}-t\right)$.

$$M_x(t)=\frac{1}{\beta^\alpha\;\Gamma(\alpha)\left(\frac{1}{\beta}-1\right)}\int_0^\infty y^{\alpha-1}\;e^{-y}dy$$

$$=\frac{\Gamma(\alpha)}{\beta^\alpha\;\Gamma(\alpha)\left(\frac{1}{\beta}-t\right)^\alpha}=\frac{1}{(1-\beta t)^\alpha}$$

$$=(1-\beta t)^{-\alpha} \tag{4.53}$$

$$M_x(t)=(1-\beta t)^{-\alpha}$$

$$=1+\alpha.\beta t+\alpha\frac{(\alpha+1)}{2\,!}(\beta t)^2$$

$$+\alpha(\alpha+1)(\alpha+2)\frac{(\beta t)^3}{3\,!}+\ldots\ldots$$

This Binomial expansion of $(1-\beta t)^{-\alpha}$ gives the various raw moments as the coefficients of t, $t^2/2\,!$ $t^3/3\,!$...

Ex. 4.7.1. *The increase in sales per day (in money value) in a particular shop, after the appointment of a new sales girl, is approximately distributed as a Gamma distribution with $\alpha=2$, $\beta=2$. What is the probability that the increase in sales-tax returns on a day selected at random exceeds Rs. 100 if sales-tax is levied at the rate of 5% ?*

Sol. Let x=increase in sales on any day. X has the distribution,

$$f(x)=(1/\beta^\alpha\;\Gamma(a))x^{\alpha-1}\quad e^{-x/\beta}$$

where $\alpha=2,\ \beta=2.$

That is, $f(x)=(1/4)x\;e^{-x/2}$ for $x>0$

$=0$ elsewhere.

If the increase in sales tax is to exceed Rs. 100 the increase in sales is to exceed Rs. 2000. Hence the required probability is,

$$=P\{x>2000\}=\int_{2000}^{\infty}(1/4)x\,e^{-x/2}\,dx$$

$$=\int_{1000}^{\infty}y\,e^{-y}\,dy \text{ (by putting } x/2=y)$$

$$=-y\,e^{-y}\Big]_{1000}^{\infty}-e^{-y}\Big]_{1000}^{\infty}$$

(by integrating by parts)

$$=1000e^{-1000}+e^{-1000}=(1001)e^{-1000}.$$

Exercises

4.24. Find the moment generating function for (*a*) a rectangular distribution with parameters α and β, (*b*) an exponential distribution with parameter θ.

4.25. Determine the distributions, if possible, from the following moment generating functions :

(*a*) $M_X(t)=(1-2t)^{-3}$, (*b*) $M_X(t)=(1-t)^{-1}$,

(*c*) $M_X(t)=e^{2t+t^2}$, (*d*) $M_X(t)=e^{2t^2}$.

(See the Normal distribution also)

4.26. Integrating by parts or otherwise, show that

$$\Gamma(\alpha)=(\alpha-1)\,.\,\Gamma(\alpha-1) \text{ for } \alpha>0,$$

4.27. Show that $\Gamma(1/2)=\sqrt{\pi}$.

[**Hint.** $[\Gamma(1/2)]^2=\int_0^{\infty}\int_0^{\infty}x^{-1/2}\,y^{-1/2}\,e^{-(x+y)/2}dxdy$. Change to polar co-ordinates]

4.28. The Beta distribution is defined by the density function

$$f(x)=k\,x^{\alpha-1}\,(1-x)^{\beta-1} \text{ for } 0<x<1,\ \alpha>0,\ \beta>0$$
$$=0 \text{ elsewhere.}$$

where $k=1/B(\alpha,\beta)=\Gamma(\alpha+\beta)/\Gamma(\alpha)\,.\,\Gamma(\beta)$.

Evaluate (*a*) E(X), (*b*) Var (X), (*c*) $P\{x\geqslant 0.5\}$ when X is a Beta variate, with $\beta=1$.

4.29. The Pearson system of curves are given by the differential equation

$$\frac{1}{f(x)} \frac{d}{dx} f(x) = (d-x)/(a+bx+cx^2).$$

Show that (1) $f(x)$ is the normal density function when $b=c=0$ and $a>0$.

(2) $f(x)$ is a Gamma density function when $a=c=0$, $b>0$ and $d=-b$.

(3) $f(x)$ is a Beta density function when $a=0$, $b=-c$ and $d>1-b$.

4.30. Evaluate the first two cumulants k_1 and k_2 for a Gamma variate and show that $k_1=\text{E(X)}$ and $k_2=\text{Var (X)}$. Also obtain k_3.

4.31. If X is a Gamma variable with parameters α and β, find the probability that $x>4$ when (1) $\alpha=1$, $\beta=2$, (2) $a=2$, $\beta=4$.

4.32. Suppose that during the rainy season on a tropical island the length of a shower has an exponential distribution with the parameter $\theta=2$, time being measured in minutes. What is the probability that a shower will last for more than 3 minutes ? If a shower has already lasted for 2 minutes, what is the probability that it will last for at least one minute more ?

4.33. The sales tax returns of a salesman is exponentially distributed with the parameter $\theta=4$. What is the probability that his sale wlll exceed Rs. 10,000, assuming that sales tax is levied at the rate of 5% on the sales ?

4.34. The annual sales of wheat in millions of bushels by a wheat board is assumed to be approximately distributed as a gamma distribution with parameters $\alpha=3$, and $\beta=2$. If this wheat board has 20 million bushels of wheat in a particular year, what is the probability that it won't be able to meet the demand ?

4.8. THE NORMAL DISTRIBUTION

This is the most important distribution in present day statistical analysis. This distribution is known by several names such as, Gaussian distribution, error curve etc. The name 'Normal distribution' is rather unfortunate. This does not in any way mean that other distributions are abnormal. Data arising from a good many practical situations are seen to be approximately normally distributed. A number of distributions can be approximated to a normal distribution. In many cases a simple transformation will transform a non normal distribution to a normal distribution. Due to many such theoretical and practical reasons this distribution plays a vital role in statistical analysis. The density function for a normal distribution is given as,

$$f(x, \theta)=k\, e^{-(x-\alpha)^2/2\beta^2} \quad -\infty<x<\infty, -\infty<\alpha<\infty, \quad \beta>0 \tag{4.54}$$

where α and β are parameters and k is a constant which can be evaluated, since

$$\int_{-\infty}^{\infty} f(x, \theta)dx=1$$

Evaluation of k:

$$\int_{-\infty}^{\infty} f(x, \theta)dx = k\int_{-\infty}^{\infty} e^{-(1/2)\left(\frac{x-\alpha}{\beta}\right)^2}dx = 1$$

Put $$\frac{x-\alpha}{\beta} = y$$

$$dx = \beta\, dy \quad \text{and} \quad -\infty < y < \infty$$

$$1 = k\int_{-\infty}^{\infty} . \beta . e^{-y^2/2}dy$$

*But $e^{-y^2/2}$, is an even function of y and hence

$$1 = k . \beta . 2\int_{0}^{\infty} e^{-y^2/2}dy.$$

Put $t = y^2/2\; dt = y . dy$ and $y = (2t)^{1/2}$

$\therefore$ $$1 = k . \beta . 2\int_{0}^{\infty} (2t)^{-1/2}e^{-t}dt$$

$$= k . \beta . \sqrt{2}\int_{0}^{\infty} t^{(1/2-1)}e^{-t}dt$$

$$= k . \beta . \sqrt{2}\Gamma(1/2)$$

$$\left[\text{Since } \Gamma(\alpha) = \int_{0}^{\infty} x^{\alpha-1}e^{-x}dx \text{ and } \Gamma(\tfrac{1}{2}) = \sqrt{\pi}\right]$$

Note. Odd and even functions.

If $\psi(x) = \psi(-x)$ then $\psi(x)$ is called an even function of x,

then $$\int_{-a}^{a} \psi(x)\, dx = 2\int_{0}^{a} \psi(x)\, dx$$

If $\phi(x) = -\phi(-x)$ then $\phi(x)$ is called an odd function of x,

then $$\int_{-a}^{a} \phi(x)\, dx = 0$$

$$=k.\beta\sqrt{2\pi}$$

$$\therefore \qquad k=\frac{1}{\beta\sqrt{2\pi}} \tag{4.55}$$

$$f(x, \theta)=\frac{1}{\beta\sqrt{2\pi}}\, e^{-(1/2)\left(\frac{x-\alpha}{\beta}\right)^2}, \; -\infty<x<\infty$$

4·81. Moments.

$$\mu=E(X)=\frac{1}{\beta\sqrt{2\pi}}\int_{-\infty}^{\infty} x\, e^{-(1/2)\left(\frac{x-\alpha}{\beta}\right)^2} dx$$

Put $y=\dfrac{x-\alpha}{\beta}$, $dx=\beta\, dy$ and $-\infty<y<\infty$

$$\therefore \qquad \mu=\frac{1}{\beta\sqrt{2\pi}}\int_{-\infty}^{\infty} (\alpha+\beta\, y)\, e^{-(1/2)\, y^2}, \; -\infty<y<\infty$$

$$=\frac{\alpha}{\beta\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(1/2)y^2} dy+\frac{\beta}{\beta\sqrt{2\pi}}\int_{-\infty}^{\infty} y\, e^{-(1/2)y^2} dy \tag{4·56}$$

But $\displaystyle\int_{-\infty}^{\infty} e^{-(1/2)y^2} dy=2\int_{0}^{\infty} e^{-(1/2)y^2} dy$ [$e^{-(1/2)y^2}$ is an even function]

and $\displaystyle\int_{-\infty}^{\infty} y\, e^{-(1/2)\, y^2} dy=0$ [$y\, e^{-(1/2)\, y^2}$ is an odd function]

$$\therefore \qquad \mu=\frac{\alpha}{\sqrt{2\pi}}\, 2\int_{0}^{\infty} e^{-(1/2)y^2} dy$$

$$=\frac{\alpha}{\sqrt{2\pi}}\ \sqrt{2\pi}$$

$$\left(\int_0^\infty e^{-(1/2)\,y^2}\,dy \text{ is evaluated in section 4·8}\right)$$

$$=\alpha \tag{4·57}$$

$$\sigma^2=\mu_2=\text{Var. }(X)=E[X-E(X)]^2=E[X-\alpha]^2$$

$$=\frac{1}{\beta\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-\alpha)^2\, e^{-(1/2)\left(\frac{x-\alpha}{\beta}\right)^2}dx$$

Put $y=\dfrac{x-\alpha}{\beta}$, then

$$\sigma^2=\frac{\beta^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2\, e^{-(1/2)\,y^2}\,dy$$

$$=\frac{\beta^2}{\sqrt{2\pi}}\ 2\int_0^\infty y^2\, e^{-(1/2)\,y^2}\,dy$$

Put $t=(1/2)\,y^2$, then

$$\sigma^2=\frac{\beta^2}{\sqrt{2\pi}}\,2\sqrt{2}\int_0^\infty y^{(3/2)-1}\, e^{-t}\,dt$$

$$=\frac{\beta^2}{\sqrt{\pi}}\ 2\Gamma(3/2)=\frac{\beta^2}{\sqrt{\pi}}\ 2\ \tfrac{1}{2}\ \Gamma\ (1/2)$$

[Since $\Gamma\ (\alpha)=(\alpha-1)\Gamma\ (\alpha-1)$ and $\Gamma\ (1/2)=\sqrt{\pi}$]

$$=\beta^2 \tag{4·58}$$

For the normal distribution the parameters α and β are such that α is $E(X)$ and β is the standard deviation of X. Because of this property a normal distribution is usually given by the density function

$$f(x,\ \theta)=\frac{1}{\sigma\sqrt{2\pi}}\ e^{-(1/2)\left(\frac{x-\mu}{\sigma}\right)^2},\quad \begin{array}{l}-\infty<x<\infty\\ -\infty<\mu<\infty,\ \sigma>0\end{array} \tag{4·59}$$

where μ and σ are the parameters. This representation reminds the readers that the parameters are E(X) and standard deviation of X when X is a normal variate. There are several notations for a normal probability distribution. Some of them are as follows.

(1) $N(\mu, \sigma)$ (Normal distribution with mean μ and S.D. σ)

(2) $N(\mu, \sigma^2)$ (Normal distribution with mean μ and variance σ^2)

(3) $X:N(\mu, \sigma)$ (The stochastic variable X follows a Normal distribution with mean μ and S.D. σ)

For example

$X:N(0, \sigma)$ (The *s.v.* X has a Normal distribution with mean 0 and S.D. σ)

i.e. $$f(x, \theta)=\frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{x^2}{2\sigma^2}}, \quad -\infty<x<\infty,\ \sigma>0 \tag{4.60}$$

$X:N(0, 1)$ (The *s.v.* X has a normal distribution with mean 0 and S.D. unity)

i.e. $$f(x)=\frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}, \quad -\infty<x<\infty \tag{4.61}$$

This is also called the standard normal distribution since $f(x)$ here is the density function of a standardized normal variate.

In this book we will use the motion $N(\mu, \sigma)$.

4·82. The Moment Generating Function. For simplicity and convenience we will find out the

M.G.F. for $Y=X-\mu$

$$M_Y(t)=M_{X-\mu}(t)=e^{-t\mu}\, M_X(t)$$

or $$M_X(t)=e^{t\mu}\, M_{X-\mu}(t)$$

$$M_{X-\mu}(t)=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{t(x-\mu)}\, e^{-(1/2)\left(\frac{x-\mu}{\sigma}\right)^2}$$

Put $x-\mu=y$ then $-\infty<y<\infty$

$$M_{X-\mu}(t)=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{ty-(1/2)\frac{y^2}{\sigma^2}}\, dy$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-(1/2)\,\sigma^2\,[y^2-2\sigma^2 t\,y]}\,dy$$

To complete the square in the exponent we add and subtract

$$\frac{1}{2\sigma^2}\,[\sigma^4\,t^2]$$

$$\therefore\quad M_{X-\mu}(t)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[y^2-2\sigma^2 ty+\sigma^4t^2]+\frac{t^2\sigma^2}{2}}\,dy$$

$$=\frac{e^{\frac{t^2\sigma^2}{2}}}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[y-\sigma^2t]^2}\,dy$$

Put
$$u=\frac{y-\sigma^2 t}{\sigma}$$

$$du=\frac{dy}{\sigma}\quad\text{and}\quad -\infty<u<\infty$$

$$M_{X-\mu}(t)=\frac{e^{t^2\sigma^2/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}}\,du$$

$$=e^{t^2\sigma^2/2}. \tag{4·63}$$

(Since $\int_{-\infty}^{\infty} e^{-u^2/2}\,du=\sqrt{2\pi}$, see exercise 4·27 at the end of this section and section 4.8)

$\therefore$ The M.G.F. $M_X(t)=e^{t\mu}M_{X-\mu}(t)$

$$=e^{t\mu+\frac{t^2\sigma^2}{2}} \tag{4·64}$$

Cor. 1. $$M_{X-\mu}(t)=e^{t^2\sigma^2/2} \tag{4·65}$$

Cor. 2. $$M_{\left(\frac{X-\mu}{\sigma}\right)}(t)=e^{t^2/2} \tag{4·66}$$

i.e. The M.G.F. of the standardized normal variate is $e^{t^2/2}$. The various raw moments may be obtained from the M.G.F. by the formula

$$\frac{d^r}{dt^r} M_X(t)\Big|_{t=0} = \mu'_r$$

4.83. Graphical Representations. Graphical representations of the distributions

(1) $X : N(\mu, \sigma)$ (2) $X : N(0, \sigma)$ (3) $X : N(0, 1)$

are given in Fig. 4.5.

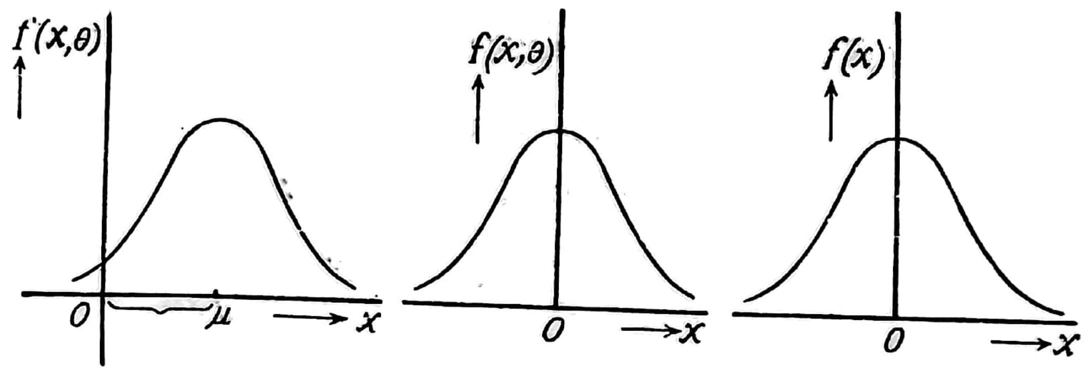


Fig. 4·5.

$f(x, \theta) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\left(\frac{1}{2}\right)\left(\frac{x-\mu}{\sigma}\right)^2}$	$f(x, \theta) = \frac{1}{\sigma.\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma^2}}$	$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$
Maximum ordinate $= \frac{1}{\sigma\sqrt{2\pi}}$.	Maximum ordinate $= \frac{1}{\sigma\sqrt{2\pi}}$.	Maximum ordinate $= \frac{1}{\sqrt{2\pi}}$.
Symmetric about the ordinate at $x = \mu$. Points of inflection at $x = \mu \pm \sigma$.	Points of inflection at $x = \pm\sigma$. Symmetric about the $f(x, 0)$ axis.	Points of inflection at $x = \pm 1$. Symmetric about the $f(x)$ axis.

Let us examine some probability statements. Consider the following problems :

Ex. 4.83.1. (a) $P\left\{\left|\frac{x-\mu}{\sigma}\right| \leqslant 1\right\} = ?$

(b) $P\left\{\left|\frac{x-\mu}{\sigma}\right| \leqslant 2\right\} = ?$

(c) $P\left\{\left|\frac{x-\mu}{\sigma}\right| \leqslant 3\right\} = ?$

where $X : N(\mu, \sigma)$.

Sol. (*a*) $P\left\{\left|\frac{x-\mu}{\sigma}\right| \leqslant 1\right\} = P\{|x-\mu| \leqslant \sigma\}$

$$= P\{-\sigma \leqslant x-\mu \leqslant \sigma\}$$

$$= P\{\mu-\sigma \leqslant x \leqslant \mu+\sigma\}$$

$$= \int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= 0{\cdot}6826 \text{ (See the comments)}$$

This is given by the shaded area in Fig. 4·6 (*a*) (4·67)

(*b*) $P\left\{\left|\frac{x-\mu}{\sigma}\right| \leqslant 2\right\} = P\{|x-\mu| \leqslant 2\sigma\}$

$$= P\{-2\sigma \leqslant x-\mu \leqslant 2\sigma\}$$

$$= P\{\mu-2\sigma \leqslant x \leqslant \mu+2\sigma\}$$

$$= \int_{\mu-2\sigma}^{\mu+2\sigma} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= 0{\cdot}9544 \qquad (4{\cdot}68)$$

This probability is given by the shaded area in Fig. 4·6 (*b*).

(*c*) $P\{|x-\mu| \leqslant 3\} = P\{\mu-3\sigma \leqslant x \leqslant \mu+3\sigma\}$

$$= \int_{\mu-3\sigma}^{\mu+3\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= 0{\cdot}9974 \qquad (4{\cdot}69)$$

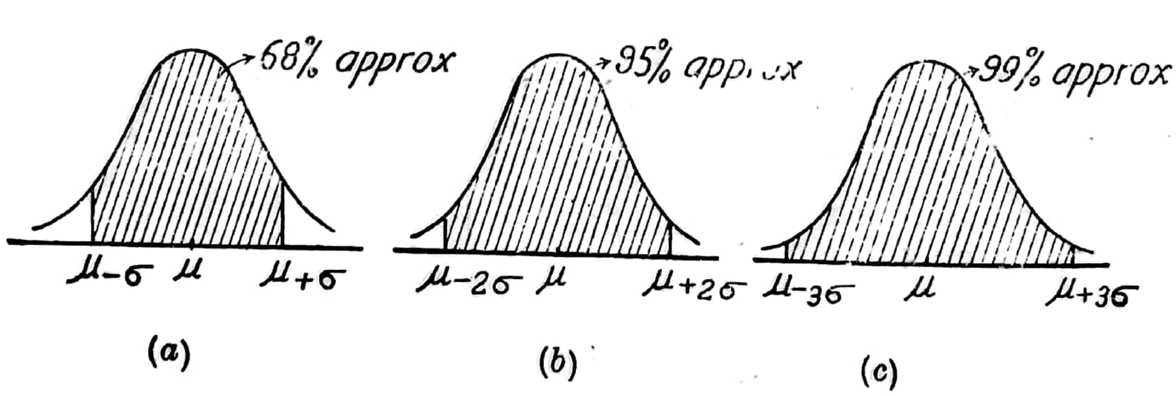


Fig. 4.6.

Comments. These areas may be found by using a normal probability table. This topic will be discussed in the next section. Prababilities that $\left|\frac{x-\mu}{\sigma}\right| \geqslant 1, 2, 3$ may be obtained by subtracting

the areas in Fig. 4·5 (*a*), (*b*), (*c*), from unity respectively.

or $$P\left\{\left|\frac{x-\mu}{\sigma}\right|\geqslant d\right\}=1-P\left\{\left|\frac{x-\mu}{\sigma}\right|<d\right\}$$

where d is any given constant.

The normal distribution X : N(0, σ) and X : N(0, 1) may be studied as special cases of X : N(μ, σ). For example N(0, σ) is N(μ, σ) when $\mu=0$ etc.

4.84. Normal Probability Tables. If we are interested in the probability that a normal variable X, distributed as N(μ, σ), takes on values greater than or equal to a given quantity c, then it may be evaluated as follows

$$P\{x\geqslant c\}=\int_c^\infty \frac{1}{\sqrt{2\pi}\sigma}\, e^{-(1/2)\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Put $$\frac{x-\mu}{\sigma}=y \text{ then}$$

when $$x=c,\ y=\frac{c-\mu}{\sigma} \text{ and } \frac{c-\mu}{\sigma}<y<\infty$$

$$\therefore \qquad P\{x\geqslant c\}=\int_t^\infty \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, dy \qquad (4\cdot70)$$

where $$t=\frac{c-\mu}{\sigma}.$$

The values of this integral

$$\left(\int_t^\infty \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, dy\right)$$

for various values of t are tabulated. Such tables are called the normal probability tables. In some tables only

$$\int_0^t \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2} dy \text{ is given.}$$

But $$\int_0^t \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, dy=0\cdot5-\int_t^\infty \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, dy. \qquad (4.71)$$

This follows from the symmetry of the distribution.

By using normal probability tables the tail areas

$$\int_t^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

may be easily determined. A normal probability table is given at the end of this book.

Ex. 4.8.2. *If $X : N(\mu=10, \sigma=2)$ find the probability that x lies between -3 and 12.*

Sol. $$P\{-3 \leqslant x \leqslant 12\} = \int_{-3}^{12} \frac{1}{2\sqrt{2\pi}} \cdot e^{-(1/2)\frac{(x-10)^2}{4}} dx$$

$$= \int_{\frac{-3-10}{2}}^{\frac{12-10}{2}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

$$\left(\text{By putting } y = \frac{x-10}{2}\right)$$

$$= \int_{-6\cdot5}^{1} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

$$= \int_{-6\cdot5}^{0} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy + \int_0^1 \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

$$= \int_0^{6\cdot5} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy + \int_0^1 \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

$$\left(\int_{-6\cdot5}^{0} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \int_0^{6\cdot5} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \text{ from symmetry}\right)$$

$$= 0\cdot5 + 0\cdot3413 = 0\cdot8413$$

(obtained from normal probability tables)

Ex. 4.8.3. *The marks for a particular subject, obtained by the students of a university in an examination, is assumed to be approximately normally distributed with mean 70 and standard deviation 5. A student taking this subject is chosen at random. What is the probability that his marks is over 80 ?*

Sol. Let X be the stochastic variable denoting the marks of any student in the set of students under consideration, then

$$X : N(\mu=70, \sigma=5).$$

$\therefore$ The required probability

$$=\int_{80}^{\infty} \frac{1}{\sqrt{2\pi}.5} e^{-(1/2)\frac{(x-70)^2}{25}}$$

$$=\int_{\frac{80-70}{5}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy=\int_{2}^{\infty}\frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

$$=0{\cdot}5-\int_{0}^{2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy,$$

$=0{\cdot}5-0{\cdot}4772=0{\cdot}0228$ (from normal tables)

Ex. 4·8·4. *From the following moment generating functions determine the corresponding probability distributions.*

(*a*) $M_X(t)=e^{2t+2t^2}$, (*b*) $M_X(t)=e^{8t^2}$

Sol. (*a*) The M.G.F. for a Normal distribution with parameters μ and σ is given by

$$M_X(t)=e^{t\mu+t^2\sigma^2/2}$$

$\therefore$ From the uniqueness property of M G.F.

$$e^{t\mu+t^2\sigma^2/2}=e^{2t+2t^2}=>\mu=2 \text{ and } \sigma=2$$

$\therefore$ The corresponding probability distribution is

N(μ, σ) where $\mu=2$ and $\sigma=2$.

(*b*) $M_X(t)=e^{8t^2}$

For a N(μ, σ), $M_X(t)=e^{t\mu+t^2\sigma^2/2}$

Comparing the expressions e^{8t^2} and $e^{t\mu+t^2\sigma^2/2}$ and using the uniqueness property of M.G.F. the corresponding distribution is N(μ, σ) where $\mu=0$ and $\sigma=4$.

The moment generating functions of some of the most commonly used univariate distributions, are given in the following table.

Distribution	Probability function $f(x, \theta)$	M.G.F. ; $M_X(t)$
1. Binomial	$\binom{N}{x} p^x q^{N-x}$ for $x=0, 1,\ldots N$ and 0 elsewhere	$(q+pe^t)^N$
2. Poisson	$\frac{\lambda^x}{x\,!} e^{-\lambda}$ for $x=0, 1, 2,\ldots$ and zero elsewhere	$e^{\lambda(e^t-1)}$
3. Normal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ for $-\infty<x<\infty$	$e^{t\mu+t^2\sigma^2/2}$
4. Exponential	$\frac{1}{\theta} e^{-x/\theta}$ for $x>0$ and zero elsewhere	$(1-\theta t)^{-1}$
5. Gamma	$\frac{1}{\beta^\alpha\,\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$ for $x>0$ and zero elsewhere	$(1-\beta t)^{-\alpha}$
6. Chi-square	$\frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}$ for $x>0$ and zero elsewhere	$(1-2t)^{-k/2}$

Exercises

4·35. Let $Y=\log X$. If Y has a normal distribution X is said to have a log-normal distribution and the density function for a log-normal distribution is given by,

$$f(x, \theta)=\frac{e^{-(\log x-\log \alpha)^2/2\beta^2}}{\beta.\sqrt{2\pi}\,.x} \text{ for } x>0,\ \alpha>0,\ \beta>0,$$

$=0$ elsewhere, where α and β are parameters. Find E(X) and Var. (X), if they exist.

4·36. A Pareto population is given by the density function,

$f(x, \theta)=p\, a^p/x^{p+1}$ for $x>a>0$ and $p>0$

$=0$ elsewhere.

Obtain E(X) and Var. (X), if they exist.

4·37. For a normal distribution $N(\mu, \sigma)$ show that $\mu_{2n+1}=0$ for $n=0, 1,\ldots$ and $\mu_4/\sigma^4=3$ where μ_r denotes the rth central moment.

4·38. For a normal distribution with the parameters $\mu=10$, $\sigma=2$ calculate the following probabilities.

(*a*) the probability that for the normal variate X, x is greater than 2·5, (*b*) the probability that x lies between -5 and 5, (*c*) the probability that $|x|<15$, (*d*) the probability that $|x|\geqslant 2$, (*e*) the probability that x is less than 3.

4·39. For a normal distribution with parameters μ, and $\sigma=1$ find t such that

(*a*) $P\{-t\leqslant x-\mu\leqslant t\}=0{\cdot}95$

(*b*) $P\{x-t\leqslant\mu\leqslant x+t\}=0{\cdot}99$

4·40. For a normal distribution with the parameters μ and σ evaluate t such that

$$P\left\{-t\leqslant\frac{x-\mu}{\sigma}\leqslant t\right\}$$
$$=P\{x-t\sigma\leqslant\mu\leqslant x+t\sigma\}=0{\cdot}95.$$

Also obtain two values t_0 and t_1 such that

that $$P\left\{-t_0\leqslant\frac{x-\mu}{\sigma}\leqslant t_1\right\}=0{\cdot}95.$$

Are t_0 and t_1 unique ?

4·41. Suppose that the heights of Canadian Citizens of a certain age group at a particular time, are assumed to be approximately normally distributed with the parameters $\mu=66''$ and $\sigma=2''$. What is the probability of getting a person in this age group whose height is as large as 70″. Above what height can we find the tallest 1% of this set of people.

4·42. If X is a N(μ, σ) then for any given α we can evaluate a t such

that $$P\left\{-t\leqslant\frac{x-\mu}{\sigma}\leqslant t\right\}$$
$$=P\{x-t\sigma\leqslant\mu\leqslant x+t\sigma\}=1-\alpha.$$

that is, μ is said to lie in the interval $x-t\sigma$ to $x+t\sigma$ with a probability $1-\alpha$. The diameter of bullets produced in a factory is a normal variate with a standard deviation $\sigma=0{\cdot}01$ units. A bullet taken at random from this factory is found to have a diameter 2 units. Obtain an interval estimate of μ of this normal population, with a probability of 0·95.

4·43. An indicator moves from a particular point to either sides. The deviation from this point is a normal variable with parameters $\mu=0$ and $\sigma=0{\cdot}1$. What is the probability (1) of getting a deviation as large as 0·2, (2) that any deviation will lie between $-0{\cdot}01$ and 0·1 ?

4·44. For a normal distribution N (μ, σ) evaluate $P\{|x-\mu|\geqslant 2\sigma\}$. Obtain a limit for this probability by using Chebyshev's inequality and compare the probabilities.

4·45. Consider a binomial distribution with parameters N and p. Under the conditions, when $N\to\infty$ and p remains a constant, show that a standardized binomial distribution approaches a standardized normal distribution.

[**Hint.** Show that the M.G.F. of the standardized binomial variate approaches the M.G.F. of the standardized normal variate.]

4·9. CHANGE OF VARIABLE

Sometimes in statistical analysis we will be interested in the distribution of a function of a given stochastic variable. For example, if X is a normal variate what is the distribution of X^2? If X is a Gamma variate what is the distribution of 2X+3? etc. Sometimes we will need the function $\phi(X)$ such that $Y=\phi(X)$ has a normal distribution where X is any given stochastic variable. In order to answer these problems, we will discuss the theory of change of variables in this section.

Theorem 4·1. If $y=\phi(x)$ is differentiable and either a monotonic increasing or decreasing function of x, then the density function $f_1(y)$ for Y is given by

$$f_1(y)=f_2(x)\left|\frac{dx}{dy}\right| \qquad \frac{dy}{dx}\neq 0 \tag{4·72}$$

where $f_2(x)$ is the density function for the stochastic variable X. We will prove the theorem when $\phi(x)$ is an increasing function of x and the proof when $\phi(x)$ is a decreasing function of x is left to the reader. Let $y=\phi(x)$ be a monotonic increasing function of x and let the curve $y=\phi(x)$ be as shown in Fig. 4·7. From Fig. 4·7, it is seen that corresponding to $x=v$ we have $y=u$ or corresponding to $y=u$ we have $x=v$. Let us take the distribution function for Y. By definition the distribution function F(y) of Y is

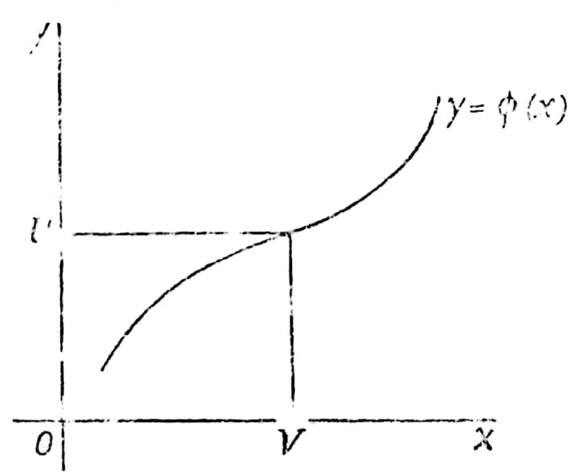


Fig. 4·7.

$$F_Y(u)=\int_{-\infty}^{u} f_1(y)\, dy=P\{y\leqslant u\}=P\{\phi(x)\leqslant u\} \tag{4·73}$$

But from Fig. 4·7 it is seen that

$$P\{\phi(x)\leqslant u\}=P\{x\leqslant v\}=\int_{-\infty}^{v} f_2(x)\,dx=F_X(v) \text{ say} \tag{4.74}$$

These results hold good for any two corresponding values of u and v.

$$\therefore \quad F_Y(u)=F_X(v) \text{ for any corresponding } u \text{ and } v. \tag{4.75}$$

$$\therefore \quad \frac{d}{du}F_Y(u)=\frac{d}{du}F_X(v)=\frac{d}{dv}F_X(v)\frac{dv}{du} \tag{4.76}$$

If $F(z)$ is the distribution function of a *s.v.* Z then the density function

$$f(z)=\frac{d}{az}F(z)$$

Equation (4.76) yields,

$$f_1(y)=f_2(x)\frac{dx}{dy} \quad \text{(since } u \text{ and } v \text{ are any corresponding values of } y \text{ and } x \text{ respectively)} \tag{4.77}$$

When $y=\phi(x)$ is a decreasing function of x, $\frac{dy}{dx}$ is negative and hence the general result may be written as

$$f_1(y)=f_2(x)\left|\frac{dx}{dy}\right| \tag{4.78}$$

Ex. 4·9·1. *If X has the density function*

$$f(x,\theta)=\frac{1}{\theta} \text{ for } 0<x<\theta,\ \theta>0$$

$$=0 \text{ elsewhere}$$

Find the density function for $Y=2X+3$.

Sol.

$$y=2x+3$$

$$\frac{dy}{dx}=2 \Rightarrow \frac{dx}{dy}=1/2$$

$$f_1(y)=f_2(x)\left|\frac{dx}{dy}\right|$$

$$f_1(y)=\frac{1}{2\theta} \text{ for } 3<y<2\theta+3.$$

$$=0 \text{ elsewhere.}$$

Ex. 4·9·2. *Given* $f(x)=x\, e^{-x^2/2}$ *for* $x>0$.
$=0$ *elsewhere*

Obtain the density function of (a) $Y=X^2$, *(b)* $Y=\log_e X$.

Sol. $y=x^2$

$$\frac{dy}{dx}=2x=2y^{1/2} \Rightarrow \frac{dx}{dy}=\tfrac{1}{2}y^{-1/2}$$

$$f_1(y)=f_2(x)\left|\frac{dx}{dy}\right|=y^{1/2}\, e^{-y/2}\,(\tfrac{1}{2}y^{-1/2})$$

$$=\tfrac{1}{2}e^{-y/2} \text{ for } 0<y<\infty$$

$$=0 \text{ elsewhere}$$

(b) $y=\log_e x$

$$\frac{dy}{dx}=\frac{1}{x} \Rightarrow \frac{dx}{dy}=x=e^y$$

$$f_1(y)=f_2(x)\left|\frac{dx}{dy}\right|$$

$$=e^{2y-(1/2)e^{2y}}$$

But $y=\log_e x \Rightarrow e^y=x$.

When $x\to 0$, $y\to-\infty$ and when $x\to\infty$, $y\to\infty$.

$$\therefore \quad f_1(y)=e^{2y-(1/2)e^{2y}}, \; -\infty<y<\infty.$$

Exercises

4.46. If X is a standardized normal variable show that X^2 is a Gamma variable with parameters $\alpha=1/2$ and $\beta=2$.

4.47. If X is a beta variable with parameters $\alpha=m/2$, $\beta=n/2$ show that $Y=nX/m(1-X)$ has an F-distribution with parameters m and n. (The F-distribution is given in section 4·12).

4.48. If X is a standardized normal variate obtain the distribution of $|X|$.

4.49. If $f(x)$ is the density function of a stochastic variable X, show that $y=\int_{-\infty}^{x} f(x)dx$ has a rectangular distribution. The change of variable here is often called the probability integral transformation.

4.50. Given that X has a density function

$$f(x)=2x/3 \quad \text{for } 0<x<1$$
$$=(3-x)/3 \quad \text{for } 1\leqslant x<3$$
$$=0 \text{ elsewhere.}$$

Obtain the distribution of $Y=X^2+1$.

4.51. (Laplace Transform). If $M_X(t)$ is the M.G.F. of a stochastic variable X, then under some general conditions, the probability distribution of X is given by

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} M_X(t)e^{-tx}dt.$$

Obtain the density functions from the following M.G.F's by using the above equation (a) $M_X(t)=e^{t^2/2}$, (b) $(1-2t)^{-3/2}$. (c) $(1-t)^{-5}$, (d) $e^{2(e^t-1)}$.

4.52. (Fourier Transform) If $\phi_x(t)$ is the characteristic function of a stochastic variable X, under some general conditions, the probability function of X is given by,

$$f(x)=(1/2\pi)\int_{-\infty}^{\infty} e^{-ixt}\phi_x(t)dt$$

where $i+(-1)^{1/2}$ and the characteristic function $\phi_x(t)$ of X is $\phi_x(t)=Ee^{itx}$ (E denotes mathematical expectation). Given the characteristic function

(a) $(2e^t/3+1/3)^5$,

(b) $pe^{it}/(1-qe^{it})$ where $0<p<1$ and $q=1-p$,

(c) $(e^{itb}-e^{ita})/it(b-a)$, (d) e^{2it-t^2},

obtain the corresponding probability distributions by using the above formula.

4.53. Show that a Binomial probability function can be written as $f(x, \theta)=\binom{n}{x}\theta^x/(1+\theta)^n$ for $x=0.1......n$, $0<\theta<\infty$, n—a positive integer.

4.54. A discrete *s.x.* is said to have a power series distribution if the probability function is given as $f(x, \theta)=a(x)\theta^x/g(\theta)$ where $a(x)$ is non-negative function of x, $g(\theta)=\sum_{x=0}^{\infty} a(x)\theta^x$ and $x=0, 1, 2,...$ By assuming special functional forms for $g(\theta)$ obtain the following distributions,

(1) Poisson, (2) Binomial, (3) Negative Binomial. (4) Geometric, (5) Logarithmic series distribution.

[**Hint :** If $g(\theta)=e^\theta$ then

$$\sum_{x=0}^{\infty} a(x)\theta^x=e^\theta = \sum_{x=0}^{\infty} \theta^x/x\,!$$

We get the Poisson distribution. If the probability function of a *s.v.* X is given by $f(x, \theta)=k \cdot \theta^x/x$, for $x=1, 2, \ldots\ldots$ then X is said to have a Logarithmic series distribution, where k is an appropriate constant.

4·55. If the probability function of a *s.v.* X is given by

$$f(x, \theta)=b(x)e^{\theta x}/h(\theta) \text{ where } x \in \text{R (real line)},$$

$$h(\theta)=\int b(x)e^{\theta x}\,dx$$

if X is continuous and is $\sum b(x)e^{\theta x}$ if X is discrete, then X is said to have a general exponential type distribution. For the general exponential type, show that $\mu(\theta)=E(X)=h'(\theta)/h(\theta)$ where $h'(\theta)=\frac{d}{d\theta}\,h(\theta)$. Also obtain as special cases the (1) Power series (thereby all the distributions in 4.54), (2) Exponential, (3) Normal with known variance, (4) Gamma with α known distributions.

BIBLIOGRAPHY

1. H.D. **Brunk.** An Introduction to Mathematical Statistics. Blaisdell Publishing Co. New York, 1965.

2. D.A.S. **Fraser**, Statistics, An Introduction. John Wiley, N.Y. 1958.

3. J.E. **Freund**, Mathematical Statistics. Prentice Hall Inc. N.J. 1964.

4. P. G. **Hoel**, Introduction to Mathematical Statistics. Wiley, N.Y. 1964.

5. M.G. **Kendall** and A. **Stuart**, The Advanced Theory of Statistics, Vol I., Charles Griffin Co. London, 1958.

6. A.M. **Mood** and F.A. **Graybill**, Introduction to the Theory of Statistics. McGraw Hill, Book Co., N.Y. 1963.

7. H.M. **Walker** and J. **Lev**, Statistical Inference. Holt Rinehart and Winston, 1953.

STATISTICAL TABLES

8. D.V. **Lindley** and J.C.P. **Miller**, Cambridge Statistical Tables. Cambridge University Press, 1958.

9. E.C. **Molina**, Poisson's Exponential Binomial Limit. D. Van Nostrand and Co. N.Y. 1943.

10. D.B. **Owen**, Handbook of Statistical Tables. Addison-Wesley, Publishing Co. Reading, Mass. 1962.

11. K. **Pearson**, Tables of the Incomplete B-functions. Cambridge University Press.

12. K. **Pearson**, Tables of the Incomplete Gamma Functions. Cambridge University Press.

13. E.S. **Pearson** and H. O. **Hartley**, Biometrika Tables for Statisticians. Vol. I. Cambridge University Press, 1954.

14. H.G. **Roming**, 50-100 Binomial Tables. Wiley, N.Y. 1953.

15. U.S. **Government.** Tables of the Binomial Probability Distributions. National Bureau of Standard, Applied Mathematics Series No. 6, Washington D.C., U.S. Govt. Printing Office, 1950.

CHAPTER 5

MULTIVARIATE DISTRIBUTIONS

5.0. Introduction. In chapters 3 and 4 we discussed probability distributions involving only one stochastic variable or univariate distributions. In this chapter we will discuss probability distributions involving more than one stochastic variable or multivariate probability distributions. If the stochastic variables are discrete we will call their joint probability distribution as a multivariate discrete distribution and if the stochastic variables are continuous then their joint distribution is called a multivariate continuous distribution. If X_1 and X_2 are two stochastic variables then their joint probability distribution is called a bivariate distribution. Similarly the joint probability distribution of k stochastic variables $X_1, X_2, \ldots X_k$ is called a k-variate distribution.

In this chapter also we will follow similar notations used in the previous chapters. The joint probability function of k stochastic variables will be denoted by $f(x_1, x_2, \ldots x_k, \theta)$ where θ stands for all the parameters in the distribution. The probability function for one stochastic variable X_1 will be denoted by $f(x_1, \theta)$, the joint probability function of two variables X_1 and X_2 will be denoted by $f(x_1, x_2, \theta)$ etc. If there is no parameter in a distribution, θ will be absent in our notation.

5.1. A BIVARIATE DISTRIBUTION

In order to introduce the concepts of joint probability distribution, the marginal probability distribution or marginal distribution and conditional distribution, we will consider an example of a bivariate discrete distribution. Let us consider a simple experiment of throwing an unbiased coin twice. The outcome set,

$$S=\{(H, H), (H, T), (T, H), (T, T)\}$$

where H and T denote 'head' and 'tail' respectively. These outcomes may be assigned probability 1/4 each.

i.e.,
$$P\{(H, H)\}=1/4$$
$$P\{(T, H)\}=1/4$$
$$P\{(H,T)\}=1/4$$
$$P\{(T, T)\}=1/4.$$

Let us consider the stochastic variables X and Y where X =number of heads and Y=number of tails then $x=0, 1, 2$ and $y=0, 1, 2$ and the distributions of X and Y are

$$
\begin{aligned}
X : f(x) &= 1/4 \text{ for } x=0 \\
&= 2/4 \text{ for } x=1 \\
&= 1/4 \text{ for } x=2 \\
&= 0 \text{ elsewhere}
\end{aligned}
$$

$$
\begin{aligned}
Y : g(y) &= 1/4 \text{ for } y=0 \\
&= 2/4 \text{ for } y=1 \\
&= 1/4 \text{ for } y=2 \\
&= 0 \text{ elsewhere.}
\end{aligned}
$$

5.11. Joint Distribution. For the example in 5·1 we will evaluate the joint probability function by forming a two-way table as given below.

Table 5.1

$y \backslash x$	0	1	2	
0	0	0	1/4	1/4
1	0	1/2	0	1/2
2	1/4	0	0	1/4
	1/4	1/2	1/4	

If the joint probability function of X and Y is denoted by $f(x, y)$ then

$$
\begin{aligned}
f(x, y) &= 1/4 \quad \text{for } x=2,\ y=0 \\
&= 1/2 \quad \text{for } x=1,\ y=1 \\
&= 1/4 \quad \text{for } x=0,\ y=2 \\
&= 0 \text{ elsewhere.}
\end{aligned}
$$

Therefore the joint distribution may be given as,

$$f(0, 2)=1/4$$
$$f(1, 1)=1/2$$
$$f(2, 0)=1/4$$

and $\quad f(x, y)=0$ elsewhere.

5.12. Marginal Distribution. In table 5.1 the probabilities in the margins, *i.e.*, $\sum_x f(x, y)$ and $\sum_y f(x, y)$ give the probability distributions of the stochastic variables Y and X respectively, where $\sum_x$ and $\sum_y$ denote summations with respect to x and y respectively. Therefore the marginal distribution may be defined as follows. If $f(x, y)$ is the joint probability function of the stochastic variables X and Y then,

$$f(x)=\sum_y f(x, y), \quad \text{if X and Y are discrete}$$

$$=\int_y f(x, y)\, dy, \quad \text{if X and Y are continuous} \tag{5.1}$$

is called the marginal distribution of X. This is the distribution of X alone. Similarly the marginal distribution of Y is given as

$$g(y)=\sum_x f(x, y), \quad \text{if X and Y are discrete} \tag{5.2}$$

$$=\int_x f(x, y)\, dx, \text{ if X and Y are continuous}$$

Here $f(x)$ and $g(y)$ need not have the same functional form.

This definition of marginal distribution may be generalized. If $f(x_1, x_2, \ldots, x_k)$ denotes the joint probability function of the stochastic variables $X_1, X_2, \ldots, X_k$ then the various marginal distribuctions are given below

$$f(x_1)=\sum_{x_2} \sum_{x_3} \ldots \sum_{x_k} f(x_1, x_2, \ldots x_k), \text{ if } X_1, X_2, \ldots, X_k \text{ are discrete} \tag{5.3}$$

$$=\int_{x_2} \int_{x_3} \cdots \int_{x_k} f(x_1, x_2, \ldots, x_k) dx_2, \ldots, dx_k$$

if $X_1, X_2, \ldots, X_k$ are continuous

$$f(x_1, x_2)= \sum_{x_3} \ldots \sum_{x_k} f(x_1, \ldots x_k), \text{ if } X_1, X_2, \ldots X_k) \text{ are discrete}$$

$$=\int_{x_3} \cdots \int_{x_k} f(x_1, x_2, \ldots, x_k)\; dx_3, \ldots, dx_k,$$

if $X_1, \ldots, X_k$ are continuous

etc. (5·4)

where $f(x_1, x_2)$ denotes the joint probability function of X_1 and X_2. Analogous to the axioms for a probability function in a univariate case we will formulate the axioms for a probability function in a multivariate case as follows. A function $f(x_1, x_2, \ldots, x_k)$ satisfying the following conditions is a probability function.

1. $f(x_1, \ldots, x_k) \geqslant 0$ for all x_i, $i=1, 2, \ldots, k$

(i.e. for all x_i, $-\infty < x_i < \infty$, $i=1, 2, \ldots, k$)

2. $$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_k)\, dx_1 \ldots dx_k = 1, \text{ if } X_1, \ldots, X_k \text{ are continuous,}$$

$$\sum_{-\infty < x_1 < \infty} \quad \sum_{-\infty < x_k < \infty} f(x_1, \ldots, x_k) = 1$$

if $X_1, \ldots, X_k$ are discrete.

In this case if A is an event $A = \{(x_1, \ldots, x_k) \mid 0 < x_1 < a_1, \ldots 0 < x_k < a_k\}$ then the probability of the occurrence of A, that is, P(A) is given by

$$P(A) = \int_0^{a_1} \cdots \int_0^{a_k} f(x_1, \ldots, x_k)\, dx_1 \ldots dx_k, \text{ if } X_1, \ldots, X_k \text{ are continuous}$$

$$= \sum_{0 < x_1 < a_1} \cdots \sum_{0 < x_k < a_k} f(x_1, \ldots, x_k)$$

if $X_1 \ldots, X_k$ are discrete.

Ex. 5·12·1. *Given that*

$$f(x, y) = \begin{cases} k\, e^{-x-2y} \text{ for } x > 0,\ y > 0 \text{ and } k, \text{ a constant,} \\ 0 \text{ elsewhere,} \end{cases}$$

is a density function. Obtain the marginal distributions or the individual density functions of X and Y.

Sol. $f(x, y)$ is a density function and, therefore,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1.$$

That is $$\int_0^{\infty} \int_0^{\infty} k\, e^{-x-2y}\, dx\, dy = k \int_0^{\infty} e^{-x}\, dx \int_0^{\infty} e^{-2y}\, dy = 1.$$

Hence $k(1/2) = 1$ or $k = 2$.

Therefore, $f(x, y)=\begin{cases} 2e^{-x-2y} & \text{for } x>0 \text{ and } y>0 \\ 0 & \text{elsewhere.} \end{cases}$

The marginal distribution of X is,

$$f(x)=\int_0^\infty 2e^{-x-2y}\,dy=e^{-x}\int_0^\infty 2e^{-2y}\,dy=e^{-x}.$$

Therefore,

$$f(x)=\begin{cases} e^{-x} & \text{for } x>0 \\ 0 & \text{elsewhere} \end{cases}$$

The marginal distribution of Y, that is, $g(y)$ is, given as,

$$g(y)=\int_0^\infty 2\,e^{-x-2y}\,dx=2\,e^{-2y}\int_0^\infty e^{-x}\,dx=2\,e^{-2y}$$

That is, $g(y)=\begin{cases} 2\,e^{-2y} & \text{for } y>0 \\ 0 & \text{elsewhere.} \end{cases}$

Ex. 5.12.2. *The joint distribution of X and Y is given below :*

$f(0, 1)=1/27$ $\quad f(0, 2)=5/27$ $\quad f(0, 3)=6/27$

$f(1, 1)=2/27$ $\quad f(1, 2)=4/27$ $\quad f(1, 3)=4/27$

$f(2, 1)=1/27$ $\quad f(2, 2)=2/27$ $\quad f(2, 3)=2/27$

and $f(x, y)$ is zero elsewhere.

Obtain the marginal distributions of X and Y.

Sol. Here $x=0, 1, 2$ and $y=1, 2, 3$.

$$f(x)=\sum_y f(x, y)$$

$=1/27+5/27+6/27=12/27$ when $x=0$

$=2/27+4/27+4/27=10/27$ when $x=1$

$=1/27+2/27+2/27=5/27$ when $x=2$.

$\therefore$ $f(x)=12/27$ for $x=0$

$=10/27$ for $x=1$

$=5/27$ for $x=2$

$=0$ elsewhere.

$$g(y)=\sum_x f(x, y)$$

$=1/27+2/27+1/27=4/27$ for $y=1$

$=5/27+4/27+2/27=11/27$ for $y=2$

$=6/27+4/27+2/27=12/27$ for $y=3$

$$\therefore \quad g(y) = \begin{cases} 4/27 & \text{for } y=1 \\ 11/27 & \text{for } y=2 \\ 12/27 & \text{for } y=3 \\ 0 & \text{elsewhere.} \end{cases}$$

A graphical representation of the joint distribution is given in Fig. 5·1.

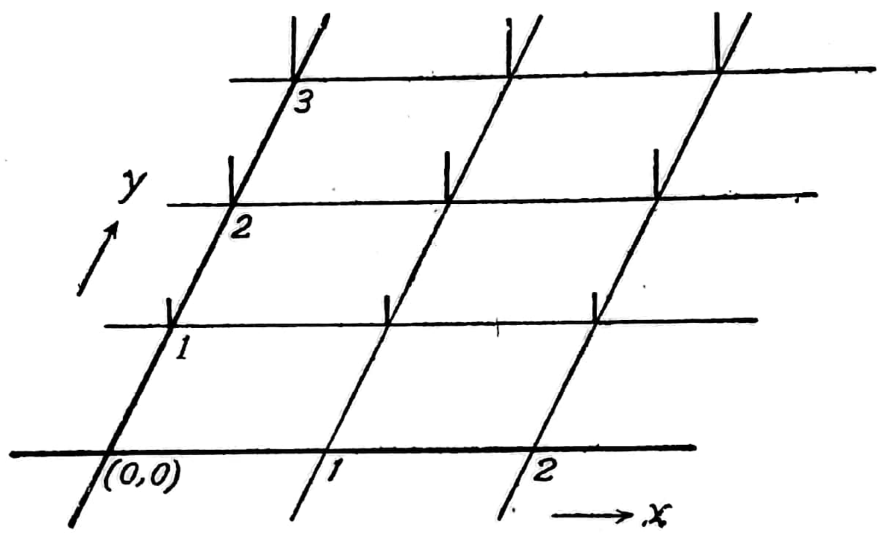


Fig. 5·1.

Comments. The marginal probability function of X is the probability distribution along the x-axis in the sense that if the total mass of one unit is to be distributed along the x-axis the masses at $x=0$, 1 and 2 give the distribution of X. Here corresponding to $x=0$ we have the masses 1/27, 5/27, and 6/27 [that is $f(0, 1)$, $f(0, 2)$ and $f(0, 3)$]. Hence the total mass at $x=0$ will be $1/27+5/27+6/27=12/27$. Similarly the masses at $x=1$ will be 10/27 and at $x=2$ will be 5/27 and thus the total is unity. If the total mass of unity is distributed along the y-axis we get the marginal probability function of Y. If we want the distribution of X along the line $y=2$, that is, if a mass of one unit is to be distributed along the line $y=2$ in the proportion of the probabilities along $y=2$ this is given by the conditional distribution of X given $y=2$. This will be discussed in the next section.

5·13. Conditional Distributions. If $f(x, y)$ is the joint probability function of two stochastic variables X and Y then the conditional distribution of X given that $y=a$, is defined as,

$$f(x \mid y=a) = \frac{f(x, y)}{g(y)} \mid y=a,$$

$$= \frac{f(x, y)}{\sum\limits_{x} f(x, y)} \mid y=a$$

for $\qquad g(a) \neq 0$, when X and Y are discrete,

$$=\frac{f(x, y)}{\int_x f(x, y)\, dx}\Bigg|_{y=a}$$

$$\text{for } g(a)=\int_x f(x, y)dx\Big|_{y=a}\neq 0$$

when X and Y are continuous. (5·5)

The conditional distribution of X given Y is the ratio of the joint distribution of X and Y to the marginal distribution of Y, evaluated at the given value of y. Similarly

$$g(y \mid x=b)=\frac{f(x, y)}{f(x)}\Bigg|_{x=b} \quad \text{if } f(b)\neq 0. \tag{5·6}$$

These definitions may be generalized. If $f(x_1, x_2,\ldots x_k)$ denotes the joint probability function of the stochastic variables $X_1, X_2,\ldots X_k$ the various conditional distributions may be given as

$$f(x_1 \mid x_2, x_3,\ldots x_k)=\frac{f(x_1, x_2,\ldots x_k)}{f(x_2, x_3,\ldots x_k)} \text{ if } f(x_2, x_3,\ldots x_k)\neq 0 \tag{5·7}$$

$$f(x_1, x_2 \mid x_3,\ldots x_k)=\frac{f(x_1, x_2,\ldots x_k)}{f(x_3, x_4\ldots x_k)} \text{ if } f(x_3,x_4,\ldots x_k)\neq 0 \tag{5·8}$$

etc.

$f(x|y)$ [is read as the probability function of X given Y ; x/y is not a division but is only a notation).

Ex. 5·13·1. *For the example 5·12·2 obtain the conditional distributions of* (*a*) *X given that $y=2$,* (*b*) *Y given that $x=0$.*

Sol. (*a*) When $y=2$, $f(x, y)$ is given as $f(0, 2)=5/27$, $f(1, 2)=4/27$, $f(2, 2)=2/27$ and $f(x, 2)=0$ elsewhere.

$g(y)$ at $y=2$ is $11/27$ (evaluated in Ex. 5·12·2)

$$\therefore \quad f(x|y=2)=\frac{f(x, y)}{g(y)}\Bigg|_{y=2}$$

$$=(5/27)/(11/27)=5/11 \text{ for } x=0$$
$$=(4/27)/(11/27)=4/11 \text{ for } x=1$$
$$=(2/27)/(11/27)=2/11 \text{ for } x=2$$
$$=0 \text{ elsewhere.}$$

$$(b)\quad g(y \mid x=0)=\frac{f(x, y)}{f(x)}\Bigg|_{x=0}$$

$$f(0, y)=1/27 \text{ for } y=1$$
$$=5/27 \text{ for } y=2$$
$$=6/27 \text{ for } y=3$$
$$=0 \text{ elsewhere.}$$

$$f(x)\,\Big|_{x=0}=12/27$$

$$\therefore \quad g(y \mid x=0)=(1/27)/(12/27)=1/12 \text{ for } y=1$$
$$=(5/27)/(12/27)=5/12 \text{ for } y=2$$
$$=(6/27)/(12/27)=6/12 \text{ for } y=3$$
$$=0 \text{ elsewhere.}$$

$g(y \mid x=0)$ denotes the conditional distribution of Y given $x=0$.

Ex. 5·13·2. *Given the joint density function*

$$f(x, y)=\tfrac{2}{3}(x+1)e^{-y},\ 0<x<1,\ y>0$$
$$=0 \text{ elsewhere.}$$

Obtain the conditional distributions of (a) x given that $y=1$, (b) Y given that $x=1/3$.

Sol. $$f(x)=\int_0^\infty \tfrac{2}{3}(x+1)e^{-y}dy=\tfrac{2}{3}(x+1)\int_0^\infty e^{-y}dy$$
$$=\tfrac{2}{3}(x+1),\ 0<x<1$$
$$=0 \text{ elsewhere.}$$

$$g(y)=\int_0^1 \tfrac{2}{3}(x+1)e^{-y}dx=e^{-y}\int_0^1 \tfrac{2}{3}(x+1)dx$$
$$=e^{-y},\quad y>0$$
$$=0 \text{ elsewhere.}$$

$$\text{(a) } f(x \mid y=1)=\frac{f(x, y)}{g(y)}\ \Big|\ y=1=\frac{\tfrac{2}{3}(x+1)e^{-1}}{e^{-1}}=2(x+1)/3.$$

That is, $$f(x \mid y=1)=\begin{cases} 2(x+1)/3 \text{ for } 0<x<1 \\ 0 \text{ elsewhere.}\end{cases}$$

$$\text{(b) } g(y \mid x=1/3)=\frac{f(x, y)}{f(x)}\ \Big|\ x=1/3=\frac{e^{-y}.8/9}{8/9}=e^{-y}$$

That is, $$g(y \mid x=1/3)=\begin{cases} e^{-y} \text{ for } y>0 \\ 0 \text{ elsewhere.}\end{cases}$$

Comments. Here it may be noticed that the conditional distribution of X is the same as the marginal distribution of X whatever may be the condition imposed on Y; and similarly for the distribution of Y. This is due to the fact that X and Y are independent. Independence of stochastic variables will be discussed later.

We will introduce one more illustration to bring out the ideas of marginal and conditional distributions. Suppose that a

person is throwing darts at a target. Let the target be the center of a 12×10 rectangular board, as shown in Fig. 5·2.

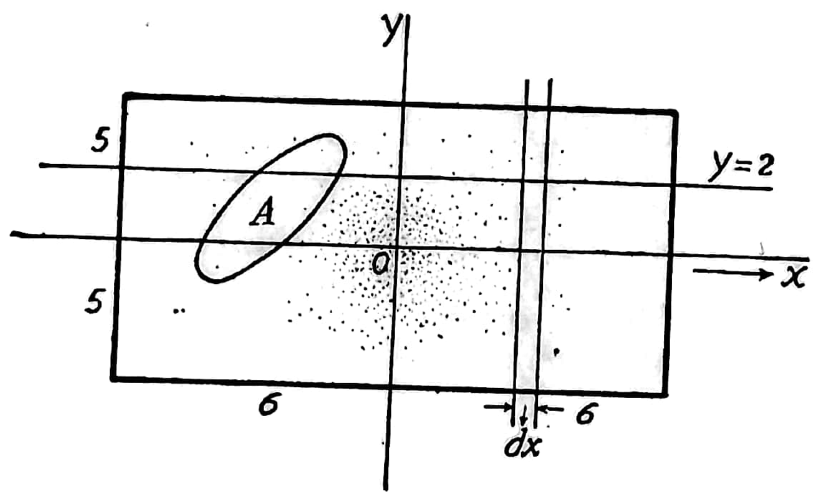


Fig. 5·2.

Let us assume that even if he misses the target he will at least hit the board. If we take a rectangular co-ordinate system as shown in Fig. 5·2, any point of hit can be represented by a pair of numbers (x, y). If he has thrown the dart for a very large number of times then a good approximation to the probability that he will hit the region A (see Fig. 5·2) is given by the relative frequency (ratio of the total number of points in A to the total number of trials). This board in Fig. 5·2 is given by the rectangle $\{(x, y) \mid -6 \leqslant x \leqslant 6, -5 \leqslant y \leqslant 5\}$. If this rectangle is divided into small squares, say of one square unit area each, by drawing lines parallel to the x and y-axes and if square pillars of heights equal to the frequencies (number of points in the corresponding squares) are erected over the squares, we get a two dimensional histogram. If this histogram is smoothed by a surface we get a surface of the form $Z = f(x, y)$.

Then, for example, the probability that the dart will hit the region, $B = \{(x, y) \mid 0 < x < 2, 0 < y < 2\}$ is proportional to

$$\int_0^2 \int_0^2 f(x, y)\, dx\, dy.$$

If the total volume under the surface is assumed to be one unit then

$$P(B) = \int_0^2 \int_0^2 f(x, y)\, dx\, dy$$

and

$$\int_{-6}^{6} \int_{-5}^{5} f(x, y)\, dx\, dy = 1 = \int_{\Omega} \int f(x, y) dx\, dy$$

where, $\Omega(\text{omega})=\{(x, y) \mid -6 \leqslant x \leqslant 6, -5 \leqslant y \leqslant 5\}$.

Ω is the sure event because we assumed that he will at least hit the board.

If we consider two *s.v's* X and Y which assume the values x and y then the joint density function of X and Y is given by

$$f(x, y) \text{ for } -6 \leqslant x \leqslant 6, \ -5 \leqslant y \leqslant 5$$

and $$f(x, y)=0 \text{ elsewhere.}$$

If we want the probability function $f(x)$ of X, that is, if we want the probability distribution of the x-co-ordinates of the points of hit, this is given by, $\int_{-5}^{5} f(x, y)dy$ since we have already assumed that the probability surface is $Z=f(x, y)$. From the experimental data we can get an approximation to the curve $f(x)$ as follows. Divide the interval $(-6, 6)$ into a number of small sub-intervals. Obtain the frequencies (number of points whose x-co-ordinates fall in the various subintervals) and draw a histogram by erecting rectangles whose areas are proportional to these frequencies and smooth this histogram by a curve. If we want the distribution of all points of hit whose y co-ordinates are 2 (or the distribution of the points of hit along the line $y=2$ as shown in Fig. 5·2) this is known as the conditional distribution of X given $y=2$ and is denoted by $g(x \mid y=2)$.

5.14. Distribution Function. If $f(x_1, x_2 \ldots\ldots x_k)$ is the joint probability function of the stochastic variables $X_1, X_2 \ldots\ldots X_k$ then the distribution function on the cumulative probability function $F(a_1, a_2, \ldots\ldots, a_k)$ is defined as,

$$F(a_1, a_2, \ldots\ldots a_k) = \sum_{-\infty < x_1 \leqslant a_1} \ \sum_{-\infty < x_2 \leqslant a_2} \cdots \sum_{-\infty < x_k \leqslant a_k} f(x_1, x_2, \ldots x_k)$$

If $X_1, X_2, \ldots\ldots X_k$ are discrete

$$= \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \cdots\cdots \int_{-\infty}^{a_k} f(x_1, x_2 \ldots x_k) dx_1, dx_2, \ldots\ldots dx_k \qquad (5.9)$$

if $X_1, X_2, \ldots\ldots X_k$ are continuous.

It may be noticed that $F(\infty, \infty, \ldots\ldots, \infty)=1$ analogous to $F(\infty)=1$ in the univariate case. $F(a_1, \ldots a_k)=0$ if any $a_i=-\infty$.

Ex. 5.14.1. *For the following probability distribution*

$$f(1, 1)=1/8, f(2, 1)=3/8$$
$$f(1, 2)=2/8, f(2, 2)=2/8$$

and $f(x, y)=0$ elsewhere, find (a) $F(0, 2)$, (b) $F(1, 3)$, (c) $F(3, 5)$.

Sol. (*a*) $$F(0, 2) = \sum_{-\infty < x \leqslant 0} \ \sum_{-\infty < y \leqslant 2} f(x, y)=0$$

(Since $f(x, y)=0$ except for $x=1, 2$ and $y=1, 2$)

(*b*) $$F(1, 3)=f(1, 1)+f(1, 2)+f(1, 3)$$
$$=1/8+2/8+0=3/8.$$

(c) $$F(3, 5)=f(1, 1)+f(1, 2)+f(2, 1)+f(2, 2)=1.$$

Ex. 5.14.2. *For the following probability distribution*

$$f(x, y, z)=x \cdot e^{-y-z/2}, \quad 0<x<1, y>0\ z>0$$
$$=0 \text{ elsewhere},$$

find $F(a, b, c)$, where a, b, c are constants >0.

Sol. $$F(a, b, c)=\int_{-\infty}^{a}\int_{-\infty}^{b}\int_{-\infty}^{c} f(x, y, z)dx\,dy\,dz$$

$$=0+\int_{0}^{a}\int_{0}^{b}\int_{0}^{c} x\,e^{-y-z/2}\,dx\,dy\,dz$$

$$=\int_{0}^{a} x\,.\,dx\int_{0}^{b} e^{-y}\,dy\int_{0}^{c} e^{-z/2}\,dz$$

$$=a^2(1-e^{-b})(1-e^{-c/2}).$$

Comments. Evidently if $a>1$, $b=\infty$ and $c=\infty$ then

$$F(a, b, c)=\int_{0}^{1}\int_{0}^{\infty}\int_{0}^{\infty} x\,e^{-y-z/2}dx\,dy\,dz=1$$

Exercises

5.1. A box contains 3 red marbles and 5 white marbles. Three marbles are taken at random with replacement. On the outcome set of this experiment define two bivariate probability distributions and obtain the respective probability functions.

5.2. Consider an experiment of rolling a balanced die twice. Let X and Y be the sum and difference rolled (*i.e.*, X=sum of the numbers appearing at the two trials etc.) Obtain the joint probability distribution of X and Y. Evaluate the distribution function. Obtain the marginal distributions of X and Y. Also, obtain the conditional distribution of X given that $y=4$.

5.3. Given the following bivariate probability distribution,

$f(-1, 0)=1/15$, $f(-1, 1)=3/15$, $f(-1, 2)=2/15$

$f(0, 0)=2/15$, $f(0, 1)=2/15$, $f(0, 2)=1/15$

$f(1, 0)=1/15$, $f(1, 1)=1/15$, $f(1, 2)=2/15$.

$f(x, y)=0$ elsewhere

obtain (1) the distribution function, (2) the marginal distributions of X and Y, (3) the conditional distribution of X given $y=2$.

5.4. G iven the distribution function,

$$F(x, y) = \begin{cases} 0 & \text{for } x<0,\ y<0 \\ 1/12 & x=0,\ y=0 \\ 3/12 & x=1,\ y=0 \\ 4/12 & x=0,\ y=1 \end{cases} = \begin{cases} 6/12 & \text{for } x=1,\ y=1 \\ 7/12 & x=0,\ y=2 \\ 1 & x\geqslant 1,\ y\geqslant 2 \end{cases}$$

obtain the marginal distribution of X and Y. Also obtain the conditional distribution of Y given $x=0$.

5.5. Given that

$$f(x, y, z) = \begin{cases} k(x+1)(y+2)(z+3), & 0<x<1,\ 0<y<1,\ 0<z<1, \\ 0 \text{ elsewhere} \end{cases}$$

and k is a constant, is a density function. Obtain (1) k, (2) the marginal distributions of X, Y and Z, (3) conditional distribution of Z given $x=a$ and $y=b$.

5.6. Given that

$$f(x, y) = \begin{cases} k(2x+3)e^{-y/2} \text{ for } 0<x<2,\ y>0 \\ 0 \text{ elsewhere, where } k \text{ is a constant,} \end{cases}$$

is a density function. Show that $f(x, y)=f(x) \cdot g(y)$.

5.7. Obtain $P\{x\geqslant 0, y\geqslant 1\}$ for the probability distributions in problems 5.3 and 5.6. Verify that

$$P\{x\geqslant 0, y\geqslant 1\}=1-P\{x<0, y<1\}.$$

5.8. Given the density function

$$f(x, y, z)=8xyz \text{ for } 0<x<1,\ 0<y<1,\ 0<z<1$$
$$=0 \text{ elsewhere,}$$

obtain $P\{0<x<0.5,\ 0.5<y<1,\ 0.5<z<0.75\}$.

5.9. Given the density function of X and the conditional density function of Y given X as

$$f(x)=\tfrac{2}{3}(x+1) \quad \text{for} \quad 0<x<1$$
$$= 0 \text{ elsewhere}$$

and

$$g(y, x) = \begin{cases} x\, e^{-xy} & \text{for } y>0 \\ 0 \text{ elsewhere,} \end{cases}$$

obtain $P\{y\geqslant 2\}$.

5.10. Given the density function of Y and the conditional density function of X given Y as,

$$g(y) = \begin{cases} 1/2 & \text{for } 0<y<2 \\ 0 \text{ elsewhere} \end{cases}$$

and
$$f(x \mid y) = \begin{cases} y\, e^{-xy} & \text{for } x > 0 \\ 0 & \text{elsewhere,} \end{cases}$$

respectively, obtain $P\{1 < 2x < 10\}$.

5.2. MOMENTS

The rth and sth product moment about the origin of two stochastic variables X and Y is defined as,

$$\mu'_{rs} = E(X^r Y^s) \tag{5.10}$$
$$= \sum_x \sum_y x^r y^s \ f(x, y),$$

if X and Y are discrete.

$$= \int_x \int_y x^r y^s f(x, y) dx\, dy,$$

if X and Y are continuous, where E denotes 'expectation'.

The r^{th} and s^{th} central moment of X and Y is defined as

$$\mu_{rs} = E[X - \mu_x]^r [Y - \mu_y]^s \tag{5.11}$$
$$= \sum_x \sum_y (x - \mu_x)^r (y - \mu_y)^s f(x, y),$$

if X and Y are discrete.

$$= \int_x \int_y (x - \mu_x)^r (y - \mu_y)^s f(x, y)\, dx\, dy,$$

if X and Y are continuous, where $\mu_x = E(X)$ and $\mu_y = E(Y)$. These definitions can be generalized for a k-variate case.

5.21. Covariance

$$\mu_{11} = E(X - \mu_x)(Y - \mu_y) \tag{5.12}$$

μ_{11} is defined as the covariance between X and Y. Evidently when Y is the same as X, the covariance becomes the variance of X. Other notations for the covariance between X and Y are C(X, Y) or C(Y, X), σxy, Cov (X, Y) etc. It may be easily seen that

$$\sigma_{xy} = E(X - \mu_x)(Y - \mu_y) = E(X.Y) - \mu_x \cdot \mu_y \tag{5.13}$$

$$= E(X \cdot Y) - E(X) \cdot E(Y).$$

Ex. 5.21.1. *For the probability distribution in Ex. 5.14.1, find the covariance between X and Y.*

Sol. Here $\quad f(x) = 3/8$ for $x = 1$
$\qquad = 5/8$ for $x = 2$
$\qquad = 0$ elsewhere

and
$$g(y)=4/8 \text{ for } y=1$$
$$=4/8 \text{ for } y=2$$
$$=0 \text{ elsewhere.}$$

$\therefore$
$$E(X)=1.(3/8)+2.(5/8)=13/8$$
$$E(Y)=1.(4/8)+2.(4/8)=12/8=3/2$$
$$E(X.Y)=(1.1)(1/8)+(1.2)(2/8)+(2.1)(3/8)+(2.2)(2/8)$$
$$=1.(1/8)+2.(2/8)+2.(3/8)+4.(2/8)=19/8.$$

(Here a dot means a multiplication).

$$\therefore \quad \sigma_{xy}=E(X.\,Y)-E(X).E(Y)=\frac{19}{8}-\left(\frac{13}{8}\right)\left(\frac{3}{2}\right)$$
$$=-1/16.$$

Comments. It was seen that the standard deviation is a measure of dispersion in a univariate distribution. Similarly the covariance of X and Y may be taken as a measure of joint dispersion in a bivariate population defined by the joint probability function of X and Y. If there is a high probability that large values of x go with large values of y and small values of x go with small values of y then the covariance may be expected to be positive ; if not, the covariance may be negative or zero. When X and Y are independent $\sigma_{xy}=0$ and this will be discussed later.

5·22. Correlation Coefficient. The correlation between the stochastic variables X and Y may be defined as

$$\rho \text{ or } \rho_{xy}=\frac{\sigma_{xy}}{\sigma_x .\ \sigma_y}=\frac{E[X-E(X)][Y-E(Y)]}{\{E[X-E(X)]^2\,E[Y-E(Y)]^2\}^{\frac{1}{2}}}$$

if σ_x and $\sigma_y \neq 0$, where ρ (rho) denotes the correlation coefficient, and σ_x and σ_y denote the standard deviations of X and Y respectively. It may be noticed that ρ_{xy} is a pure coefficient, *i.e.*, it is independent of the units of measurements, scaling etc., of X and Y since we divide the covariance by the standard deviations of X and Y. ρ_{xy} is a measure of the relationship between X and Y and plays an important role in correlation analysis in Statistics. It can be shown that $-1 \leqslant \rho \leqslant 1$. This correlation coefficient is called a linear correlation coefficient or a simple product moment correlation coefficient. There are other types of correlation coefficients such as partial correlation, multiple correlation, serial correlation, biserial correlation, curve linear correlation etc., which will not be discussed here. For further reading see the references given at the end of this chapter.

Ex. 5·22·1. *Two stochastic variables X and Y are given as $Y=aX+b$ where a and b are constants, show that $|\rho|=1$ where ρ is the correlation coefficient between X and Y.*

Sol. $Y=aX+b.$

$\therefore$ $E(Y)=a.E(X)+b$ and Var. $(Y)=a^2$ Var. (X)

$$Y-E(Y)=a[X-E(X)]$$

$\therefore$ $\text{Cov }(X, Y)=E[Y-E(Y)][X-E(X)]=Ea[X-E(X)]^2$

$$=a.\sigma_1^2 \quad \text{where } \sigma_1^2=\text{Var. }(X).$$

$$=\frac{\text{Cov }(X, Y)}{\{\text{Var }(X).\ \text{Var.}(Y)\}^{\frac{1}{2}}}=\frac{a.\ \sigma_1^2}{\{\sigma_1^2\, a^2\, \sigma_1^2\}^{\frac{1}{2}}}$$

$$=\frac{a}{|a|}=\pm 1.$$

(Since the standard deviation is defined as the positive square root of the variance.)

$\therefore$ $|\rho|=1.$

Comments. It is seen that when there is a linear relationship between X and Y, $|\rho|=1$ or there is perfect correlation. In the light of this result ρ may be considered to be a measure of linear dependence between X and Y.

Ex. 5·22·2. *Given that the joint density function of X and Y as*

$$f(x, y)=\begin{cases} e^{-x-y}, & x>0,\ y>0 \\ 0 & \text{elsewhere} \end{cases}$$

evaluate the correlation between X and Y.

Sol. $$\rho=\frac{E(X-\mu_x][Y-\mu_y]}{\sigma_x\,\sigma_y}=\frac{E(XY)-E(X).E(Y)}{\sigma_x\,\sigma_y}$$

Here $E(X.Y)=\int_0^\infty \int_0^\infty xy\, e^{-x-y}\, dx\, dy$

$$=\int_0^\infty x\, e^{-x}\, dx \int_0^\infty y\, e^{-y}\, dy.$$

$$E(X).E(Y)=\int_0^\infty x\, e^{-x}\, dx \int_0^\infty y\, e^{-y}\, dy$$

(Since $f(x)=\begin{cases} e^{-x} \text{ for } x>0 \\ 0 \text{ elsewhere}\end{cases}$ and $g(y)=\begin{cases} e^{-y} \text{ for } y>0 \\ 0 \text{ elsewhere}\end{cases}$

$\therefore \quad \sigma_{xy}=E(X.Y)-E(X).E(Y)=0$

Comments. For this particular example it is seen that $\rho=0$, because $\sigma_{xy}=0$. It so happened that in our example $\rho=0$, but in general ρ need not be zero. In this example X and Y are independent and hence $\rho=0$. Independence of stochastic variables will be discussed later.

5·23. Conditional Expectation. Conditional expectation of X given Y may be denoted by E(X | Y). This gives the expectation of X in the conditional distribution of X at the given point for Y. The conditional distribution of X given $y=a$ is usually denoted by

$$f(x \mid y)=\frac{f(x, y)}{g(y)} \text{ at } y=a, \text{ if } g(a)\neq 0.$$

Therefore the conditional expectation,

$$E(X \mid Y)=\int_{-\infty}^{\infty} x.f(x \mid y)\, dx, \text{ if X and Y are continuous,}$$

$$=\sum_x x\, f(x \mid y), \text{ if X and Y are discrete,} \quad (5\cdot15)$$

In general the conditional expectation of a function of X, say $\psi(X)$ may be written as

$$E\{\psi(X) \mid Y\}=\int_x \psi(x).f(x \mid y)dx, \text{ if X and Y are continuous,} \quad (5\cdot16)$$

$$=\sum_x \psi(x)\, f(x \mid y), \text{ if X and Y are discrete.}$$

These definitions may be generalized to a general multivariate distribution. We may notice that E(X | Y) is not a *s.v.* but is only a function of y or a function of a if the condition is given as $y=a$.

Ex. 5·23·1. *For the following bivariate distribution find the expectation of* X^2 *given that* $y=1$.

$f(1, 1)=1/10, f(1, 2)=3/10$

$f(2, 1)=1/10, f(2, 2)=1/10$

$f(3, 1)=2/10, f(3, 2)=2/0$

and $f(x, y)=0$ *elsewhere.*

Sol. The marginal distribution of Y is obtained as

$$g(y)=4/10 \text{ for } y=1$$
$$=6/10 \text{ for } y=2$$
$$=0 \text{ elsewhere}$$

$\therefore$ $$g(y)=4/10 \text{ when } y=1$$

$\therefore$ $$f(x \mid y)=\frac{f(x, y)}{g(y)} \text{ at } y=1$$
$$=1/4 \text{ for } x=1$$
$$=1/4 \text{ for } x=2$$
$$=2/4 \text{ for } x=3$$
$$=0 \text{ elsewhere.}$$

$\therefore$ The conditional expectation of X^2 given that $y=1$ is

$$E(X^2 \mid Y)=1^2(1/4)+2^2(1/4)+3^2(2/4)=23/4.$$

Ex. 5·23·2. *Given the joint density of X, Y and Z as*

$$f(x, y, z)=\tfrac{1}{2}\, x\, e^{-y-z} \text{ for } 0<x<2,\ y>0,\ z>0$$
$$=0 \text{ elsewhere,}$$

find the conditional expectation of Y given X and Z.

Sol. The conditional distribution of Y given X and Z is,

$$g(y \mid x, z)=\frac{f(x, y, z)}{h(x, z)}=\frac{x\, e^{-y-z}/2}{\int_0^\infty xe^{-y-z}dy/2}$$

$$=\begin{cases} e^{-y} & \text{for } y>0 \\ 0 & \text{elsewhere,}\end{cases}$$

where $h(x, z)$ is the marginal distribution of X and Z. Therefore,

$$E(Y \mid X, Z)=\int_0^\infty y\, g(y \mid x, z)dy=\int_0^\infty ye^{-y}\, dy=1.$$

Comments. The conditional expectation of a *s.v.* X is the E(X) in the conditional distribution of X. The conditional distribution is illustrated by two examples in sections 5.1 and 5.2.

Exercises

5.11. Find μ'_{20}, μ'_{02}, μ'_{11} and μ'_{12} for the following distribution : $f(-1, 0)=1/8$, $f(-1, 1)=2/8$, $f(1, 0)=3/8$, $f(1, 1)=2/8$ and $f(x, y)=0$ elsewhere.

5.12. Obtain μ'_{20}, μ'_{02}, μ'_{11} and μ'_{12} for the following distribution :

$$f(x, y)=\begin{cases} 3x\, e^{-3y/2} & \text{for } 0<x<2,\ y>0 \\ 0 & \text{elsewhere.}\end{cases}$$

5.13. Obtain the product moment correlation coefficient between X and Y in problem 5.12.

5.14. If C is a constant and X and Y are stochastic variables, show that (1) $E(C \mid X) = C$, (2) $E(CY \mid X) = CE(Y \mid X)$.

5.15. If X, Y, Z are stochastic variables with finite means show that (1) $E[(X+Y) \mid Z] = E(X \mid Z) + E(Y \mid Z)$, (2) $E[E(Y \mid X)] = E(Y)$, (3) $\text{Var } Y = E[\text{Var}(Y \mid X)] + \text{Var}[E(Y \mid X)]$ when in (2) and (3) $E(Y \mid X)$ is treated as a function of a stochastic variable X.

5.16. Using the results in section 1.52 show that the sample correlation coefficient r is such that $-1 \leqslant r \leqslant 1$.

5.17. Using the result that the variance of a stochastic variable is never negative, prove that $-1 \leqslant \rho \leqslant 1$ by taking

$$\text{Var}\left(\frac{X}{\sigma_1} + \frac{Y}{\sigma_2}\right) \text{ and Var}\left(\frac{X}{\sigma_1} - \frac{Y}{\sigma_2}\right)$$ where $\sigma_1^2 = \text{Var}(X)$ and $\sigma_2^2 = \text{Var}(Y)$.

5.18. Show that the sample correlation coefficient r of the ranks obtained by the n students in a class for two subjects is

$$r = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$ where d_i is the difference of the ranks of the i^{th} student.

Assume that there are no ties. (**Hint :** The ranks will be numbers from 1 to n in some order.)

5.19. Obtain the rank correlation between the following ranks.

x	1	3	2	4	6	5
y	2	5	3	1	4	6

5.3. SPECIAL MULTIVARIATE PROBABILITY MODELS

In this section only a few of the important multivariate distributions will be discussed. In a univariate case we have seen that the Binomial distribution is the most important discrete distribution and that the normal distribution is the most important continuous distribution. Analogous to Binomial and univariate normal, we will discuss Multinomial and Multivariate Normal distributions in the following sections.

5.31. The Multinomial Distribution. Consider an experiment where each trial results in one of the k mutually exclusive outcomes with probabilities $p_1, p_2, \ldots \ldots p_k$, where

$$\sum_{i=1}^{k} p_i = 1.$$

(For example if a blanced die, is rolled once, one and only one of the numbers 1, 2, 3, 4, 5 or 6 is obtained. Each occurs with probability 1/6 and $1/6+1/6+...+1/6=1$). If such a trial is repeated n times what is the probability of getting exactly x_1 outcomes of the first kind, x_2 outcomes of the second kind,... and x_k outcomes of the k^{th} kind such that $x_1+x_2+...+x_k=n$? This can be easily evaluated by a procedure similar to that in the case of a Binomial distribution. The probability of getting $x_1,...x_2,...,x_k$ in some specified order is evidently $p_1^{x_1} p_2^{x_2}...p_k^{x_k}$. Therefore the probability of getting exactly x_1 outcomes of the i^{th} kind for $i=1, 2,...k$ is

$$f(x_1, x_2,...x_k^{\theta})=\frac{n\,!}{x_1\,!\,x_2\,!\,...\,x_k\,!}\,p_1^{x_1}\,p_2^{x_2}\,...p_k^{x_k} \tag{5.17}$$

This may be considered to be the joint probability of the stochastic variables $x_1, x_2 ... x_k$ where x_i is the number of outcomes of the i^{th} kind for $i=1, 2,...k$. $0 \leqslant x_i \leqslant n$ for $i=1, 2,...n$ and $x_1+x_2+...+x_k=n$, $p_1+p_2+...+p_k=1$, $p_i \geqslant 0$ for all i.

This distribution has the parameters $p_1, p_2,...p_k$ and n. But $p_1+p_2+...+p_k=1$. Therefore there are k parameters, n and $p_1, p_2,...p_{k-1}$.

It may be noticed that $\sum\limits_{x_1} \sum\limits_{x_2} ... \sum\limits_{x_k} f(x_1, x_2,...x_k, \theta)$

$$=(p_1+p_2+...+p_k)^n=1. \tag{5.18}$$

Here $X_1, X_2, ...,X_k$ are said to have a multinomial distributions with the parameters n and $p_1, p_2,...,p_k$ where $\Sigma p_i=1$.

Ex. 5.31.1. *It is given that in a community the probabilities of getting a person having black hair, brown hair, and blonde hair are 0.50, 0·30, 0.20 respectively. If 10 persons are selected at random from this community what is the probability of getting 4 having black hair, 2 having brown hair and 4 having blonde hair ?*

Sol. This may be compared to a multinomial probability situation with $n=10$, $p_1=0.50$, $p_2=0.30$, and $p_3=0.20$. Therefore the required probability is

$$f(4, 2, 4)=\frac{10\,!}{4\,!\,2\,!\,4\,!}(0.50)^4(0.30)^2(0.20)^4=0.028.$$

5.32. The Bivariate Normal Distribution. If X and Y are two continuous stochastic variables having the joint density function of the form

$$f(x, y, \theta)=\frac{1}{2\pi\ \beta_1\beta_2(1-\delta^2)^{1/2}}\times$$
$$e^{-\frac{1}{1-\delta^2}\left[\left(\frac{x-\alpha_1}{\beta_1}\right)^2-2\delta\left(\frac{x-\alpha_1}{\beta_1}\right)\left(\frac{x-\alpha_2}{\beta_2}\right)+\left(\frac{y-\alpha_2}{\beta_2}\right)^2\right]} \qquad (5.19)$$

for $-\infty<x<\infty$, $-\infty<y<\infty$, $\beta_1>0$, $\beta_2>0$, $-1<\delta<1$,

$-\infty<\alpha_1<\infty$, $-\infty<\alpha_2<\infty$,

then X and Y are said to have a Bivariate normal distribution. Here α_1, α_2, β_1, β_2 and δ are parameters. Later we will show that $E(X)=\alpha_1$, $E(Y)=\alpha_2$. $\sigma_x=\beta_1$, $\sigma_y=\beta_2$ and $\rho=\delta$.

5.32.1. Marginal Distribution of X. The marginal distribution of X is obtained as

$$f(x, \theta)=\int_{-\infty}^{\infty} f(x, y, \theta)dy$$

Put $u=\frac{x-\alpha_1}{\beta_1}$ and $v=\frac{y-\alpha_2}{\beta_2} \Rightarrow dv=\frac{dy}{\beta_2}$

$$\therefore\quad f(x, y, \theta)=\frac{1}{2\pi\beta_1\beta_2\sqrt{1-\delta^2}}\, e^{-\frac{1}{2(1-\delta^2)}[u^2-2\delta uv+v^2]} \qquad (5.20)$$

$$\therefore\quad f(x, \theta)=\frac{1}{2\pi\ \beta_1\beta_2\sqrt{1-\delta^2}}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\delta^2)}[u^2-2\delta uv+v^2]}\cdot\beta_2\cdot dv$$

$$=\beta_1\frac{e^{-\frac{u^2}{2(1-\delta^2)}}}{2\pi\sqrt{1-\delta^2}}\int_{-\infty}^{\infty} e^{-\frac{[v^2-2\delta uv]}{2(1-\delta^2)}}.dv \qquad (5.21)$$

But $\qquad v^2-2\delta uv+\delta^2u^2-\delta^2u^2=(v-\delta u)^2-\delta^2u^2$

$$\therefore\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\delta^2)}[v^2-2\delta uv]}dv$$

$$=e^{\frac{\delta^2u^2}{2(1-\delta^2)}}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\delta^2)}(v-\delta u)^2}dv \qquad (5.22)$$

Put $\qquad \frac{v-\delta u}{\sqrt{1-\delta^2}}=t \Rightarrow dv=\sqrt{1-\delta^2}\, dt.$

$$\therefore\ e^{\frac{\delta^2u^2}{2(1-\delta^2)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\delta^2)}(v-\delta u)^2} dv$$

$$=\sqrt{1-\delta^2}\ e^{\frac{\delta^2u^2}{2(1-\delta^2)}} \int_{-\infty}^{\infty} e^{-t^2/2}dt$$

$$=\sqrt{1-\delta^2}\ .\ e^{\frac{\delta^2u^2}{2(1-\delta^2)}} \sqrt{2\pi}$$

$$\therefore \qquad f(x,\theta)=\frac{e^{-\frac{u^2}{2(1-\delta^2)}}}{\beta_1 2\pi\sqrt{1-\delta^2}}\cdot \sqrt{1-\delta^2}\sqrt{2\pi}\ e^{\frac{\delta^2u^2}{2(1-\delta^2)}}$$

$$=\frac{e^{-u^2/2}}{\beta_1\sqrt{2\pi}}=\frac{1}{\beta_1\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x+\alpha_1}{\beta_1}\right)^2} \quad -\infty<x<\infty \tag{5.23}$$

i.e. $f(x, \theta)$ is a univariate normal distribution.

In the case of univariate distribution it was seen that $\alpha_1=E(X)$ and β_1=standard deviation of X.

From the symmetry of $f(x, y, \theta)$ it follows that the marginal distribution of Y is

$$g(y,\theta)=\frac{1}{\sqrt{2\pi}\ \beta_2}\ e^{-\frac{1}{2}\left(\frac{y-\alpha_2}{\beta_2}\right)^2} \quad -\infty<y<\infty \tag{5.24}$$

5.32.2. Covariance between X and Y.

$$\sigma_{xy}=E(X-\alpha_1)(Y-\alpha_2)$$

$$=\frac{1}{\beta_1\beta_2 2\pi\sqrt{1-\delta^2}}\times$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\alpha_1)(y-\alpha_2)\ e^{\frac{1}{2(1-\delta^2)}\left[\left(\frac{x-\alpha_1}{\beta_1}\right)^2-2\delta\left(\frac{x-\alpha_1}{\beta_1}\right)\left(\frac{y-\alpha_2}{\beta_2}\right)+\left(\frac{y-\alpha_2}{\beta_2}\right)^2\right]}$$

$$dx\,dy. \tag{5.25}$$

Put $\qquad u=\dfrac{x-\alpha_1}{\beta_1}$

and $\qquad v=\dfrac{y-\alpha_2}{\beta_2}\Rightarrow dx\,dy=\beta_1\beta_2\,du\ dv$

(See Section 5.5 for change of variables.)

$$\sigma_{xy}=\frac{\beta_1\beta_2}{2\pi\sqrt{1-\delta^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} uv\, e^{-\frac{1}{2(1-\delta^2)}[u^2-2\delta uv+v^2]}\, du\, dv.$$

$$=\frac{\beta_1\beta_2}{2\pi\sqrt{1-\delta^2}}\int_{-\infty}^{\infty} v\, e^{-\frac{v^2}{2(1-\delta^2)}} \cdot \left[\int_{-\infty}^{\infty} u \cdot e^{-\frac{1}{2(1-\delta^2)}[u^2-2\delta uv]}\, du\right]dv$$

$$=\frac{\beta_1\beta_2}{2\pi\sqrt{1-\delta^2}}\int_{-\infty}^{\infty} v \cdot e^{-\frac{v^2}{2(1-\delta^2)}+\frac{\delta^2 v^2}{2(1-\delta^2)}}$$

$$\left[\int_{-\infty}^{\infty} u\, e^{-\frac{1}{2(1-\delta^2)}(u-\delta v)^2}\, du\right]dv \qquad (5{\cdot}26)$$

But $\displaystyle\int_{-\infty}^{\infty} u\, e^{-\frac{1}{2(1-\delta^2)}(u-\delta v)^2}\, du=\delta . v\int_{-\infty}^{\infty} e^{-\frac{t^2}{2(1-\delta^2)}}dt$

$$=\delta \cdot v\sqrt{1-\delta^2} \cdot \sqrt{2\pi}$$

$$\therefore \quad \sigma_{xy}=\frac{\delta \cdot \beta_1 \cdot \beta_2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} v^2 \cdot e^{-\frac{v^2}{2}}\, dv=\delta \cdot \beta_1 \cdot \beta_3$$

Therefore the covariance between X and Y is $\delta \cdot \beta_1 \cdot \beta_2$, and the correlation coefficient between X and Y is

$$\rho=\frac{\sigma_{xy}}{\sigma_x\,\sigma_y}=\frac{\delta \cdot \beta_1 \cdot \beta_2}{\beta_1 \cdot \beta_2}=\delta. \qquad (5{\cdot}27)$$

Because of these properties the parameters in a Bivariate normal distribution are usually denoted as μ_1, μ_2, σ_1, σ_2 and ρ. Thus the joint density function of X and Y is written as

$$f(x, y, \theta)=\frac{1}{\sigma_1\sigma_2 2\pi\sqrt{1-\rho^2}} \exp. -\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right.$$

$$\left.-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right] \qquad (5{\cdot}28)$$

$$-\infty<x<\infty,\ -\infty<y<\infty,\ \sigma_1>0,\ \sigma_2>0 \text{ and } -1<\rho<1$$

$$-\infty<\mu_1<\infty,\ -\infty<\mu_2<\infty,$$

where μ_1, μ_2, σ_1, σ_2 and ρ are parameters and 'exp.' denotes 'exponential'. It is seen that

$\mu_1=E(X)$, $\mu_2=E(Y)$, $\sigma_1^2=\text{Var}(X)$, $\sigma_2^2=\text{Var}(Y)$

$\rho=\text{Cov}(X, Y)/\sigma_1 \cdot \sigma_2$.

5.32.3. The conditional Distribution of X given Y. The conditional distribution of X given Y is by definition,

$$f(x/y)=\frac{f(x, y)}{g(y)} \text{ if } g(y)\neq 0$$

where y is a given quantity. We have already seen that the marginal distribution of Y is $N(\mu_2, \sigma_2)$ when the joint distribution is given with the parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ and ρ.

$$\therefore\ f(x \mid y)=\frac{1}{\sigma_1\sigma_2 2\pi\sqrt{1-\rho^2}}\exp-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 -2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\cdot\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right] \quad (5.29)$$

divided by $\dfrac{1}{\sigma_2\sqrt{2\pi}}\exp\left\{-\dfrac{1}{2\sigma_2^2}\left(\dfrac{y-\mu_2}{\sigma_2}\right)^2\right\}$

This upon simplification will be easily reduced to

$$f(x \mid y)=\frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-\rho^2}}\exp.-\frac{1}{2(1-\rho)^2}\left[\frac{x-\mu_1-\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)}{\sigma_1}\right]^2$$

$-\infty < x < \infty$. Hence y is a given quantity. (5.30)

Evidently $f(x \mid y)$ is a normal distribution with mean

$=\mu_1+\rho\dfrac{\sigma_1}{\sigma_2}(y-\mu_2)$ and with standard deviation equal to $\sigma_1\sqrt{1-\rho^2}$.

Similarly it can be easily shown that $g(y \mid x)$ is a

$$N\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\ \sigma_2\sqrt{1-\rho^2}\right).$$

If $\rho=0$ evidently $f(x \mid y)=f(x)$ and $g(y \mid x)=g(y)$.

5.32.4. The Conditional Expectation. In a Bivariate normal distribution the conditional distributions of X given Y as well as that of Y given X are given in section 5.32.3. It is seen that

$f(x \mid y)$ is normal with mean

$$=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_3)$$

and with standard deviation

$$=\sigma_1\sqrt{1-\rho^2}.$$

Therefore,

$$E(X \mid Y)=\mu_1+\rho.\ \frac{\sigma_1}{\sigma_2}(y-\mu_2)$$

and
$$E(Y \mid X)=\mu_2+\rho.\ \frac{\sigma_2}{\sigma_1}(x-\mu_1) \tag{5.31}$$

Here μ_1, μ_2, ρ, σ_1 and σ_2 are constants and hence if the conditional expectation of X given Y for various values of y, is plotted, evidently we get a straight line. Similarly the conditional expectation of Y for various given values of x defines another straight line. These lines are called the normal regression lines which will be discussed in detail in the last chapter. Since the correlation coefficient $\rho=\sigma_{12}/\sigma_1\sigma_2$ where σ_{12} is the covariance between X and Y, these regression lines may be written as :

$$E(X \mid Y)=\mu_1+\sigma_{12}(y-\mu_2)/\sigma_2$$

and
$$E(Y \mid X)=\mu_2+\sigma_{12}(x-\mu_1)/\sigma_1 \tag{5.32}$$

Ex. 5·32·1. *Rockets are fired from a rocket launcher to hit a bridge 2 units long. If the mid-point of the bridge is taken as the origin and the direction of the bridge as the x-axis the distribution of the points of hit is given to be bivariate normal with the parameters $\mu_1=0=\mu_2$, $\sigma_1=2$, $\sigma_2=3$ and $\rho=0$. What is the probability that (1) two out of three rockets fired will hit the bridge, (2) a rocket fired will hit the region $-1\leqslant x\leqslant 1$, $-2\leqslant y\leqslant 2$.*

Sol. The joint probability function is given to be,

$$f(x, y)=\frac{1}{2\pi.2.3}\ e^{-(x^2/4+y^2/9)/2},\quad -\infty<x<\infty,\ -\infty<y<\infty$$

In (1) we want $\binom{3}{2}\,p^2(1-p)^1$ where p=probability of a hit

$$=P(-1\leqslant x\leqslant 1, -c\leqslant y\leqslant c)$$

where $2c$ is the breadth of the bridge.

$$P=\int_{-1}^{1}\left(\int_{-c}^{e} f(x, y)dy\right)dx=\int_{-1}^{1}\frac{1}{2\sqrt{2\pi}}\ e^{-x^2/8}\,dx.$$

$$\int_{-e}^{c}\frac{1}{3\sqrt{2\pi}}\ e^{-y^2/18}\,dy$$

$$=\left(2\int_0^{1/2} e^{-t^2/2}\,dt\,/\sqrt{2\pi}\quad \int_0^{c/3} e^{-t^2/2}\,dt\,/\sqrt{2\pi}\right)$$

$$=0{\cdot}3830\left(2\int_0^{c/3} e^{-t^2}\,dt/\sqrt{2\pi}\right)$$

In (2) we want $\quad P(-1\leqslant x\leqslant 1,\ -2\leqslant y\leqslant 2)$

$$=\int_{-1}^{1}\left(\int_{-2}^{2} f(x,\ y,)dy\right)dx$$

$$=\int_{-1}^{1} e^{-x^2/8}/2\sqrt{2\pi}\ dx.\int_{-2}^{2} e^{-y^2/18}/3\sqrt{2\pi}\ dy$$

$$=\left(2\int_0^{1/2} e^{-z^2/2}/\sqrt{2\pi}\ dz\right)\left(2\int_0^{2/3} e^{-t^2/2}\,dt\,/\sqrt{2\pi}\right)$$

$=0{\cdot}1895$ approximately.

(from normal tables).

5·32·5. The bivariate normal surface. The univariate normal curve $N(\mu,\ \sigma)$ was seen to be symmetric about the ordinate

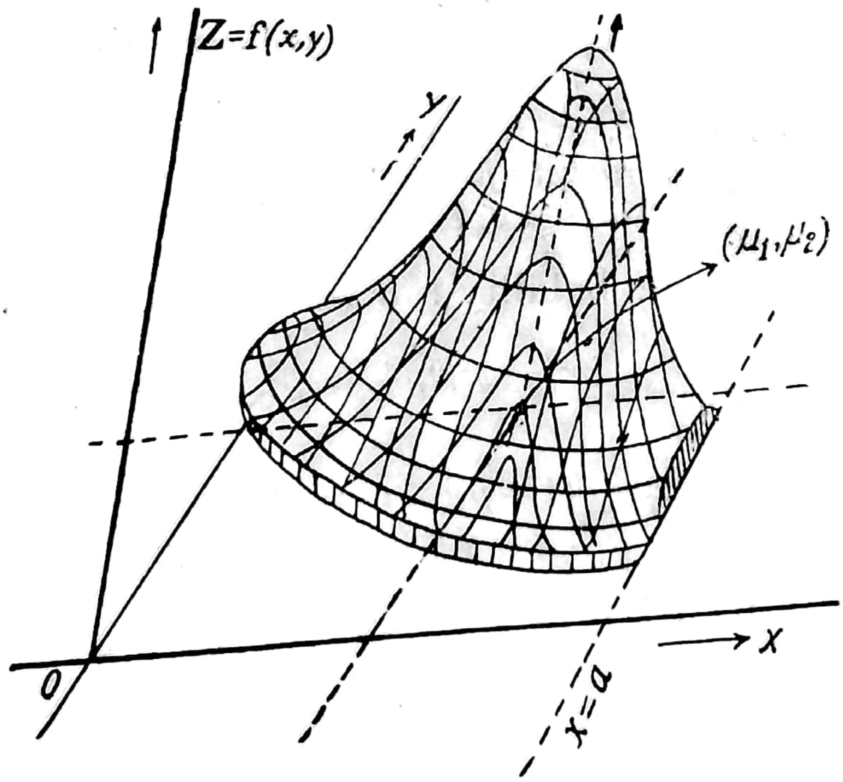


Fig. 5.2·

at $x=\mu$, having the maximum ordinate at $x=\mu$ and extending from $-\infty$ to ∞. If we plot the function $z=f(x,\ y,\ \theta)$, where $f(x,\ y,\ \theta)$ is the joint bivariate normal density function, in a three dimensional space, we get a surface in a three dimensional space. This

surface may be called the normal surface meaning the surface defined by the density function of a Bivariate normal distribution, and which has the peak at the point (μ_1, μ_2) and which is given in Fig. 5·2 and 5·3.

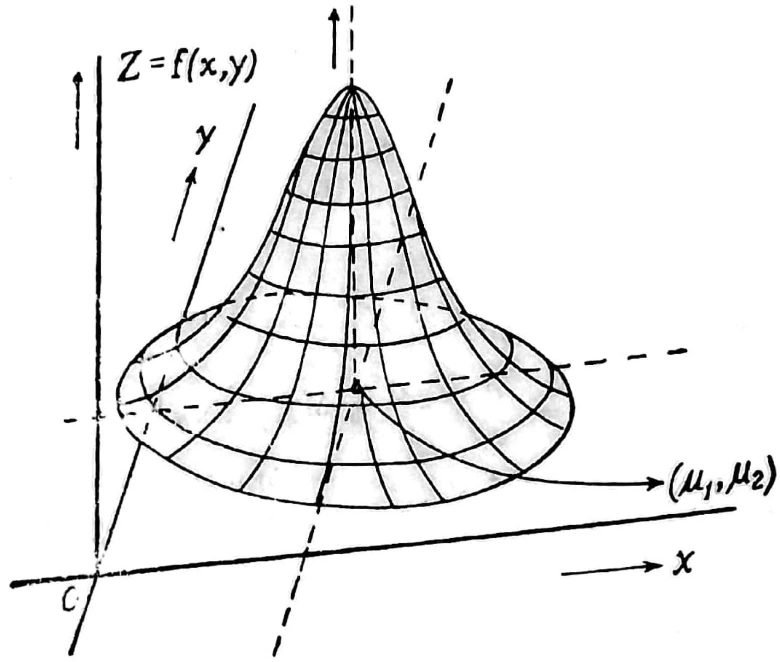


Fig. 5·3.

5·33. The Multivariate Normal Distribution. The density function in a Bivariate normal distribution can be written by using Matrix notation as follows :

$$f(x, y, \theta)=\frac{|A|^{1/2}}{(2\pi)^{2/2}} e^{-\frac{1}{2}(X-\mu)A(X-\mu)'} \tag{5·33}$$

where $\quad X-\mu=(x-\mu_1, y-\mu_2)$

is a vector of order 2. (Here X does not mean that X is a *s.v.* This is only a convenient notation here.)

$(X-\mu)'$ is the transpose of $X-\mu$.

$$A=\begin{bmatrix} \frac{1}{(1-\rho^2)\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2(1-\rho^2)} \\ \frac{-\rho}{\sigma_1\sigma_2(1-\rho^2)} & \frac{1}{\sigma_2^2(1-\rho^2)} \end{bmatrix}=\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}^{-1}$$
$$=V^{-1} \text{ (say)}$$

is a matrix of order 2 and $|A|$ is the determinant of this square matrix A. V is called the covariance matrix. The theory of matrices and determinants is given in chapter 1. A k-variate normal density function may be written as

$$f(x_1, x_2, \ldots x_k, \theta)=\frac{|A|^{1/2}}{(2\pi)^{k/2}} e^{-\frac{1}{2}(X-\mu)A(X-\mu)'}, \tag{5·34}$$

where $\quad X-\mu=(x_1-\mu_1, x_2-\mu_2, \ldots x_k-\mu_k)$

$(X-\mu)'$ is the transpose of $X-\mu$.

$A=V^{-1}$ where V is a square matrix of order k and whose $(ij)^{th}$ element is the covariance between the stochastic variables

X_i and X_j. This means that the diagonal elements of V are the variances of $X_1, X_2, \ldots X_k$ and the non-diagonal elements are the various covariances. $(X-\mu)A(X-\mu)'$, is assumed to be a positive definite quadratic form or A is assumed to be positive definite (see chapter 1)

Exercises

5·20. The moment generating function in a multivariate distribution may be defined as

$$M_{X_1 \ldots X_k}(t_1, t_2, \ldots t_k) = E\, e^{(t_1X_1 + t_2X_2 + \ldots t_kX_k)}$$

Show that (1) $E(X_i) = \dfrac{\partial}{\partial t_i} M_{X_1, \ldots X_k}(t_1, \ldots t_k)$ at $t_1 = \ldots = t_k = 0$

(2) $E(X_iX_j) = \dfrac{\partial}{\partial t_i}\dfrac{\partial}{\partial t_j} M_{X_1 \ldots X_k}(t_1, \ldots t_k)$ at $t_1 = \ldots = t_k = 0$

5·21. Obtain the moment generating function for a multinomial distribution and evaluate the product moments $\mu'_{10}, \mu'_{20}, \mu'_{02}, \mu'_{11}$ between the i^{th} and the j^{th} variables.

5·22. A bivariate normal distribution has the parameters

$\mu_1 = 30$, $\mu_2 = 20$, $\sigma_1 = 2$, $\sigma_2 = 3$ and $\rho = 0$, obtain

$P\{20 \leqslant x_1 \leqslant 35 \;\; 15 \leqslant x_2 \leqslant 25\}$.

5·23. If $f(x_1, x_2, x_3)$ is the probability function for a trinomial distribution show that the conditional distribution $f(x_1, x_2 \mid x_3)$ is a binomial distribution.

5·24. The exponent of a bivariate normal density function is

$$-\tfrac{2}{3}[(x-10)^2/4 - (x-10)(y-15)/6 + (y-15)^2/9]$$

Find out the parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ and ρ, if they exist.

5·25. Shri Chacko is shooting at a target. Let x and y be the co-ordinates of a hit, taking the center of the target as the origin. Assuming that (X, Y) has a bivariate normal distribution with the parameters,

$\mu_1 = 0 = \mu_2$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\rho = 0$, what is the probability that he will hit the target (1) if the target is a square with sides equal to 4 units, (2) the target is a circle of radius 3.

5·4. INDEPENDENCE OF STOCHASTIC VARIABLES

Two stochastic variables X and Y are said to be independent if their joint probability function is the product of the marginal probability functions of X and Y *i.e.*, according to our notation

$$f(x, y, \theta) = f(x, \theta).g(y, \theta) \tag{5·35}$$

where $f(x, \theta)$ and $g(y, \theta)$ are the marginal probability functions of X and Y respectively.

It is easily noticed that if X and Y are independent

$$f(x \mid y) = f(x) \text{ and } g(y \mid x) = g(y).$$

Ex. 5·4·1. *Check whether X and Y are independent, if the joint density function is given as*

$$f(x, y) = \tfrac{2}{5}(x+2)e^{-y},\; 0 < x < 1,\; y > 0$$
$$= 0 \text{ elsewhere.}$$

Sol. $f(x)=\int_0^{\infty} \frac{2}{5}(x+2)\, e^{-y}\, dy=\frac{2}{5}(x+2)\int_0^{\infty} e^{-y}\, dy$

$=\frac{2}{5}(x+2)$ for $0<x<1$

$=0$ elsewhere.

$g(y)=\int_0^1 \frac{2}{5}(x+2)e^{-y}\, dx=e^{-y}\int_0^1 \frac{2}{5}(x+2)\, dx$

$=e^{-y}$ for $y>0$

$=0$ elsewhere.

Therefore $f(x).\, g(y)=\frac{2}{5}(x+2)\, e^{-y}=f(x, y)$.

Hence X and Y are statistically independent.

Comments. Here it is seen that $f(x \mid y)=f(x)$
and $g(y \mid x)=g(y)$.

This definition of independence may be generalised for a number of stochastic variables. The stochastic variables $X_1, X_2, \ldots, X_k$ are said to be independent if their joint probability function equals the product of their marginal probability functions. That is,

$$f(x_1, x_2, \ldots, x_k)=f_1(x_1)\ldots f_k(x_k) \tag{5·36}$$

where $f_1(x_1), \ldots, f_k(x_k)$ are the marginal probability functions of $X_1, X_2, \ldots, X_k$ respectively.

Ex. 5·4·2. *In the following distribution (1) check for independence of X and Y, (2) obtain the conditional distribution of X given $y=\frac{1}{2}$, (3) obtain the conditional expectation of X given $y=\frac{1}{2}$. (4) obtain $P(0\leqslant x\leqslant 2 \mid y=\frac{1}{2})$.*

$$f(x, y)=\begin{cases} 2, & 0<x<1,\ 0<x<y<1 \\ 0 & \text{elsewhere.} \end{cases}$$

Sol. The region where $f(x, y)$ has non-zero values is given by the shaded area in Fig. 5.4.

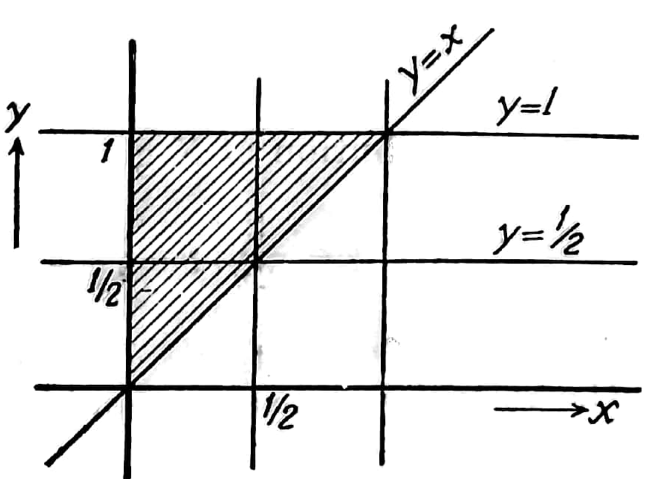


Fig. 5·4.

$$\text{Evidently} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = \int_0^1 \left(\int_0^y 2\, dx \right) dy$$

$$= \int_0^1 \left(\int_x^1 2\, dy \right) dx = 1$$

Here x varies from 0 to y and hence the marginal probability function of Y is,

$$g(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_0^y 2\, dx = 2y.$$

Along the y-axis y varies from 0 to 1 and therefore

$$g(y) = \begin{cases} 2y, & 0 < y < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Similarly the marginal distribution of X is,

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_x^1 2\, dy = 2(1-x).$$

Along the x-axis x varies from 0 to 1 and therefore,

$$f(x) = \begin{cases} 2(1-x), & 0 < x < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

(1) But $f(x).\, g(y) = 2(1-x)(2y) \neq f(x, y) = 2$ and hence X and Y are not independent.

(2) The conditional distribution of X given $y = \frac{1}{2}$ is,

$$f(x \mid y = \tfrac{1}{2}) = \frac{f(x, y)}{g(y)} \bigg|_{y=\frac{1}{2}} = \frac{2}{2y} \bigg|_{y=\frac{1}{2}} = 2.$$

But along $y = \frac{1}{2}$, x varies from 0 to $\frac{1}{2}$. Hence the distribution of X along $y = \frac{1}{2}$ or the conditional distribution of X given $y = \frac{1}{2}$ is,

$$f(x \mid y = \tfrac{1}{2}) = \begin{cases} 2, & 0 < x < \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$$

(3) $E(X \mid y=\frac{1}{2})=\int_{-\infty}^{\infty} x\, f(x \mid y=\frac{1}{2})\, dx=\int_{0}^{\frac{1}{2}} x\, 2\, dx=\frac{1}{4}.$

(4) $P(0\leqslant x\leqslant 2 \mid y=\frac{1}{2})=\int_{0}^{2} f(x \mid y=\frac{1}{2})\, dx$

$$=\int_{0}^{\frac{1}{2}} 2\, dx+\int_{\frac{1}{2}}^{2} 0\, dx=1.$$

Ex. 5.4.3. *If s.v.'s X and Y are independent, show that for any two events $A=\{x \mid a<x<b\}$ and $B=\{y \mid c<y<d\}$,*

$$P(A\cap B)=P(A).P(B).$$

Sol. Let $f(x, y)$, $f(x)$, $g(y)$ be the joint probability and marginal probability functions of X and Y respectively. Since X and Y, are independent, $f(x, y)=f(x).g(y)$. For convenience we will assume that X and Y are continuous. The discrete and mixed cases can be dealt with in a similar way.

$$P(A)=P(a<x<b)=\int_{a}^{b} f(x)\, dx$$

$$P(B)=\int_{a}^{b} g(y)\, dy \text{ and } P(A\cap B)=\int_{a}^{b}\left(\int_{c}^{d} f(x, y)\, dy\right) dx$$

$$=\int_{a}^{b}\left(\int_{c}^{d} f(x).g(y)\, dy\right) dx$$

(since X and Y are independent)

$$=\int_{a}^{b} f(x)\, dx.\int_{c}^{d} g(y)\, dy=P(A).P(B).$$

Comments. Stochastic independence of events was defined in chapter 2. Since an event can be considered to be a subset of the range of a *s.v.* all the results obtained in chapter 2 can be obtained as special cases from the general properties of some appropriate *s.v*'s.

Theorem 5.1. If X and Y are independent stochastic variables and if $\psi(X)$ and $\phi(Y)$ are functions of X and Y respectively then,

$$E[\psi(X).\phi(Y)]=E[\psi(X)].E[\phi(Y)], \qquad (5{\cdot}37)$$

where E denotes, 'mathematical expectation', and $\psi(X)$ and $\phi(Y)$ are assumed to be stochastic variables.

Proof. Case 1. Let X and Y be continuous.

$$E[\psi(X).\phi(Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x).\phi(y)\, f(x, y)\, dx\, dy$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x)\, \phi(y)\, f(x).g(y)\, dx\, dy$$

(since X and Y are independent)

$$=\int_{-\infty}^{\infty} \psi(x)\, f(x)\, dx \int_{-\infty}^{\infty} \phi(y)\, g(y)\, dy=E[\psi(X)].E[\phi(Y)].$$

The proof when X and Y are discrete, is left to the reader. It is easy to show that $E[\psi(X, Y)+\phi(X, Y)]=E\psi(X, Y)+E\phi(X, Y)$ where ψ and ϕ are functions of X and Y.

Corollary 1. The covariance between two stochastic variables X and Y is zero when X and Y are independent. [The converse, however, need not be true in general. If X and Y are normal and if Cov. (X, Y) is zero then X and Y can be shown to be independent].

Proof. $\text{Cov.}(X, Y)=E[X-E(X)][Y-E(Y)]$

$=E[X-E(X)].E[Y-E(Y)]$

(by theorem 5·1)

$=0$ (since $E[X-E(X)]=0=E[Y-E(Y)]$)

Exercises

5.26. Show that, in a bivariate normal distribution the variables are independent if and only if $\rho=0$.

5·27. Check whether the variables in the following distributions are independent.

(a) $f(x, y, z)=\begin{cases} 2\, e^{-x-y-z} & \text{for } x>0,\ y>0,\ z>0 \\ 0 & \text{elsewhere.}\end{cases}$

(b) $f(x, y)=\begin{cases} (x+1)\,(2y+1)/9 & \text{for } 0<x<1,\ 0<y<2 \\ 0 & \text{elsewhere.}\end{cases}$

5.28. Obtain the expected value and variance of (1) $Z=X+2Y$, (2) $T=X-Y$, if (a) X and Y are independent, (b) X and Y are not independent, where $E(X)=\mu_1$, $E(Y)=\mu_2$, $Var(X)=\sigma_1^2$ and $Var\ (Y)=\sigma_2^2$.

5.29. If X and Y are independent obtain the covariance between X+Y and X−Y.

5.5. CHANGE OF VARIABLES

In section 4.9 we discussed change of variables in a univariate case. Here we will consider the problem of change of variables in a multivariate distribution. If the stochastic variables $X_1, X_2, \ldots X_n$ are transformed to the stochastic variables $Y_1, Y_2, \ldots Y_n$ by the transformation

$$\begin{aligned} Y_1 &= \phi(X_1, X_2, \ldots X_n) \\ Y_2 &= \phi_2(X_1, X_2, \ldots X_n) \\ &\vdots \\ Y_n &= \phi_n(X_1, X_2, \ldots X_n) \end{aligned} \tag{5.38}$$

then the joint density function of $Y_1, Y_2, \ldots Y_n$ is given by theorem 5·2 without proof where $\phi_1, \phi_2, \ldots \phi_n$ are functions of $X_1, X_2, \ldots X_n$.

Theorem 5·2. If $X_1, X_2, \ldots X_n$ and $Y_1, Y_2, \ldots Y_n$ are as defined above then

$$f_1(x_1, x_2, \ldots x_n) = f_2(y_1, y_2, \ldots y_n) \mid J \mid \tag{5·39}$$

where $f_1(x_1, x_2, \ldots x_n)$ and $f_2(y_1, y_2, \ldots y_n)$ are the joint density functions of $X_1, X_2, \ldots X_n$ and $Y_1, Y_2, \ldots Y_n$ respectively ; $\mid J \mid$ is the absolute value of the Jacobian J where J is the following determinant

$$J = \begin{vmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_2}{\partial x_1} & \cdots & \dfrac{\partial y_n}{\partial x_1} \\ \dfrac{\partial y_1}{\partial x_2} & \dfrac{\partial y_2}{\partial x_2} & \cdots & \dfrac{\partial y_n}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial y_1}{\partial x_n} & \dfrac{\partial y_2}{\partial x_n} & \cdots & \dfrac{\partial y_n}{\partial x_n} \end{vmatrix}$$

In this theorem the differentiability of $Y_1, Y_2, \ldots Y_n$, $J \neq 0$ etc. and some general conditions on the variables are assumed.

Ex. 5·5.1. *Given the joint density function of X_1, and X_2 as*

$$f(x_1, x_2) = \tfrac{1}{2} x_1 \, e^{-x_2} \text{ for } 0 < x_1 < 2,\ x_2 > 0$$
$$= 0 \text{ elsewhere}$$

find the distribution of $X_1 + X_2$.

Sol. Let us consider a transformation

$$y_1 = x_1 + x_2 \text{ and } y_2 = x_2$$

$$J=\begin{vmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_2}{\partial x_1} \\[2ex] \dfrac{\partial y_1}{\partial x_2} & \dfrac{\partial y_2}{\partial x_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$

$\therefore \quad f_2(y_1, y_2)=f_1(x_1, x_2)=\frac{1}{2}(y_1-y_2)\,e^{-y_2}$

The region of integration is given in Fig. 5·5.

$$\left.\begin{aligned} y_1&=x_1+x_2 \\ y_2&=x_2 \end{aligned}\right\} \Rightarrow \quad \begin{aligned} x_1&=y_1-y_2 \\ x_2&=y_2 \end{aligned}$$

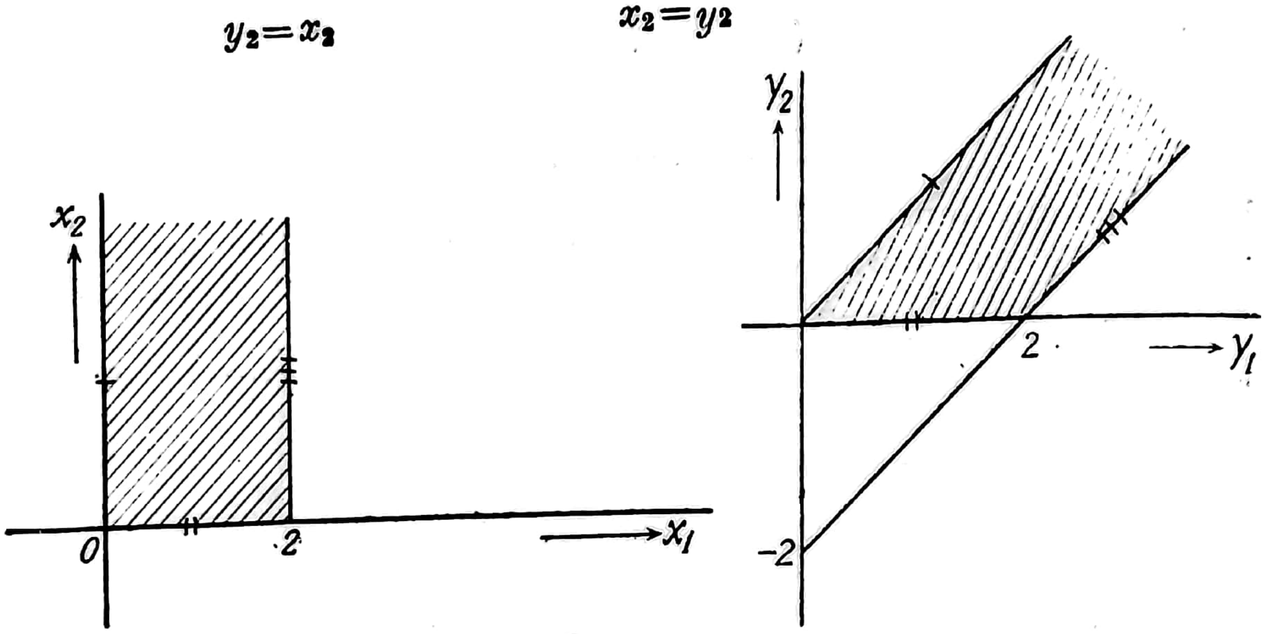


Fig. 5·5.

For $0<y_1\leqslant 2$

The density function $f(y_1)$ of Y_1 is

$$\therefore \quad f(y_1)=\int_0^{y_1} \tfrac{1}{2}(y_1-y_2)\,e^{-y_2}\,dy_2$$

$$=\tfrac{1}{2}[y_1-1+e^{-y_1}]$$

For $2<y_1<\infty$

$$\therefore \quad f(y_1)=\int_{y_1-2}^{y_1} \tfrac{1}{2}(y_1-y_2)\,e^{-y_2}\,dy_2$$

$$=\tfrac{1}{2}e^{-y_1}(1+e^2)$$

Hence the required density function is

$$f(x)=\tfrac{1}{2}\begin{cases} e^{-x}+x-1 & \text{for } 0<x\leqslant 2 \\ e^{-x}(1+e^2) & \text{for } 2<x<\infty \\ 0 \text{ elsewhere.} & \end{cases}$$

Ex. 5·5.2. *Given a bivariate rectangular distribution as*

$f(x_1, x_2)=\frac{1}{2}$ *for* $0<x_1<1$ *and* $0<x_2<2$

$=0$ *elsewhere,*

find the distribution of X_1+X_2.

Sol. Let $y_1 = x_1 + x_2$ $\Rightarrow$ $x_1 = y_2$

$y_2 = x_1$ $\quad$ $x_2 = y_1 - y_2$

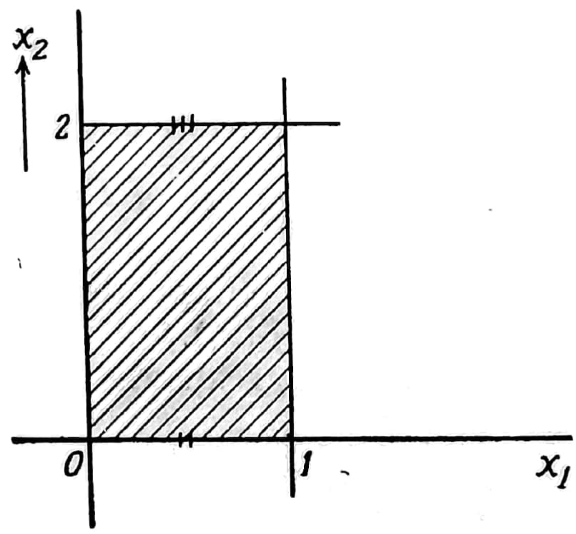
The Jacobian of the transformation is

$$J = \begin{vmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_2}{\partial x_1} \\ \dfrac{\partial y_1}{\partial x_2} & \dfrac{\partial y_2}{\partial x_2} \end{vmatrix} = -1$$

$\therefore$ $f_2(y_1, y_2) = f_1(x_1, x_2) = \frac{1}{2}$

$$\therefore \quad f(y_1) = \int_{y_2} f(y_1, y_2)\, dy_2 = \int_{y_2} \tfrac{1}{2}\, dy_2$$

The region of integration is given in Fig. 5·6.



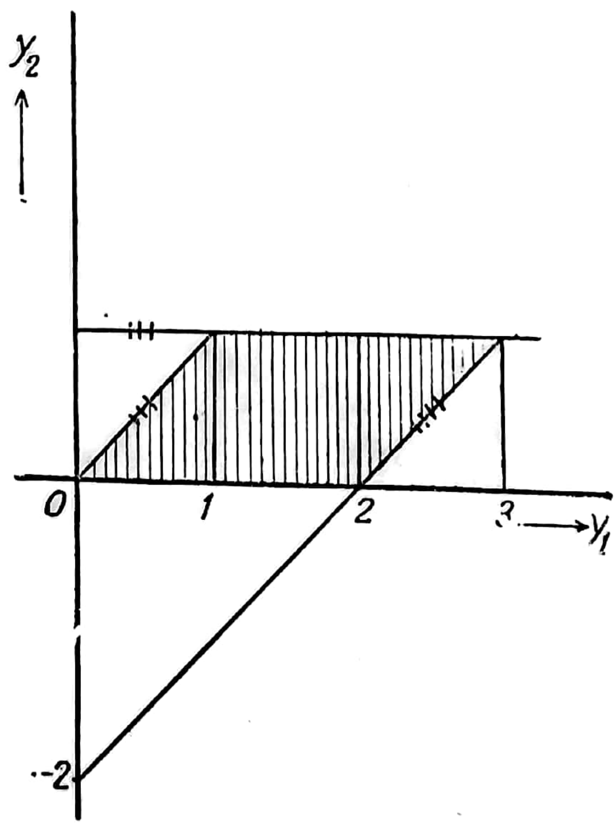


Fig. 5·6.

$$\therefore \quad f(y_1)=\int_0^{y_1} \tfrac{1}{2}\, dy_2=\left(\frac{y_1}{2}\right) \text{ for } 0<y_1<1$$

$$=\int_0^1 1/2\, dy_2=1/2 \text{ for } 1 \leqslant y_1 < 2$$

$$=\int_{y_1-2}^1 1/2\, dy_2=(1/2)(3-y_1) \text{ for } 2 \leqslant y_1 < 3.$$

Hence the density function of x_1+x_2 is

$$f(x)=\begin{cases} (1/2)(x) & \text{for } 0 < x \leqslant 1 \\ 1/2 & \text{for } 1 < x \leqslant 2 \\ (1/2)(3-x) & \text{for } 2 < x < 3 \\ 0 \text{ elsewhere.} \end{cases}$$

Ex. 5.5.3. *Two independent stochastic variables X_1 and X_2 have Gamma distributions with parameters $\alpha=m/2$, $\beta=2$ and $\alpha=n/2$, $\beta=2$ respectively. Find the distribution of* $F=\dfrac{X_1/m}{X_2/n}$.

Sol. Since X_1 and X_2 are independent, the joint density function of X_1 and X_2 is $f(x_1, x_2)=f_1(x_1).f_2(x_2)$ where f_1 and f_2 are the density functions of X_1 and X_2 respectively.

$$i.e.\quad f(x_1, x_2)=\frac{1}{2^{m/2}\Gamma\left(\frac{m}{2}\right)}\, x_1^{\frac{m}{2}-1}\, e^{-x_1/2}\, \frac{1}{2^{n/2}\Gamma(n/2)}\, x_2^{\frac{n}{2}-1}\, e^{-x_2/2}$$

$$=\frac{1}{2^{(m+n)/2}\,\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\, x_1^{\frac{m}{2}-1}\, x_2^{\frac{n}{2}-1}\, e^{-(x_1+x_2)/2} \quad \text{for } x_1>0,\ x_2>0$$

$=0$ elsewhere

Let us make the transformation

$$y_1=\frac{x_1/m}{x_2/n} \text{ and } y_2=x_2$$

$$J=\begin{vmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_2}{\partial x_1} \\ \dfrac{\partial y_1}{\partial x_2} & \dfrac{\partial y_2}{\partial x_2} \end{vmatrix} = \frac{n}{m}\,\frac{1}{x_2}$$

$$\therefore \quad g(y_1, y_2)=f(x_1, x_2)\frac{1}{|J|}=\frac{1}{2^{(m+n)/2}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}$$

$$\left(\frac{m}{n}\right)^{\frac{m}{2}-1}(y_1 y_2)^{\frac{m}{2}-1} y_2^{\frac{n}{2}-1}$$

$$e^{\frac{1}{2}\left[\frac{m}{n}y_1y_2+y_2\right]}\cdot\left(\frac{n}{m}\frac{1}{y_2}\right)^{-1}$$

$$=\frac{\left(\frac{m}{n}\right)^{\frac{m}{2}}}{2^{(m+1)/2}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}y_1^{\frac{m}{2}-1}y_2^{(m+n)/2-1}$$

$$e^{-\frac{1}{2}\left(\frac{m}{n}y_1+1\right)y_2}$$

when $g(y_1, y_2)$ is the joint density function of Y_1 and Y_2.

$$\therefore \qquad f(y_1)=\int_0^{\infty} g(y_1, y_2)dy_2$$

$$=\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{n}\right)^{\frac{m}{2}}\frac{y_1^{\frac{m}{2}-1}}{\left(1+\frac{m}{n}y_1\right)^{\frac{m+n}{2}}}$$

$$0<y_1<\infty.$$

Comments. This particular y_1 which is the ratio of two independent χ^2's divided by their corresponding degrees of freedoms is called an F-statistic and has an F-distribution given by $f(y_1)$. This distribution will be discussed in detail in the chapter on 'Sampling from Normal Distributions'.

Exercises

5.30. If (X_1, X_2) has a bivariate normal distribution with the parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ and ρ, obtain the distribution of (T_1, T_2) where,

$$T_1=(X_1-\mu_1)/\sigma_1\,;\ T_2=(1-\rho^2)^{-1}/2[(X_1-\mu_2)/\sigma_2{}^2-(X_1-\mu_1)/\sigma_1],$$

5.31. If X has a Beta distribution with the parameters $\alpha=m/2$, $\beta=n/2$, obtain the distribution of

$$Y=(n/m)\left(\frac{X}{1-X}\right).$$

Compare the distribution of Y with the F-distribution given in section 4.12.

5.32. If $X_1, X_2 \ldots X_n$ are independently and identically distributed as a Cauchy distribution (section 4.12) with $\theta=0$, show that $Y=X_1+\ldots+X_n$ has a Cauchy distribution.

5.33. If $X_1, \ldots, X_n$ are independently and identically distributed as a $N(\mu, \sigma)$ obtain the distribution of the sample mean $\bar{X}=(X_1+\ldots+X)/_n$ by using the method of change of variables.

BIBLIOGRAPHY

1. H. **Brunk**, An Introduction to Mathematical Statistics, Ginn & Co., Boston 1960.

2. H. **Cramer**, Mathematical Methods of Statistics, Princeton University Press, N.J. 1961.

3. **R. Hogg** & A. **Craig**, Introduction to Mathematical Statistics. The Macmillan Co. N.Y., 1959.

4. M. G. **Kendall** & A. **Stuart**, The Advanced Theory of Statistics, Vol. I. Charles Griffin and Co. London, 1958.

5. A.M. **Mood** & F.A. **Graybill**, Introduction to the Theory of Statistics. McGraw-Hill Book Co, N.Y. 1963.

6. G. B. **Thomas**, Calculus with Analytical Geometry, Addison Wesley, Reading Mass, 1965.

7. S.S. **Wilks**, Mathematical Statistics, John Wiley and Sons, N.Y. 1962.

TABLES OF RANDOM NUMBERS CAN BE FOUND IN THE FOLLOWING TABLES

8. D.V. **Lindley** & J.C.P. **Miller**, Cambridge Elementary Statistical Tables, Cambridge University Press, 1958.

9. D.V. **Owen**, Handbook of Statistical Tables. Addison Wesley, 1962.

STOCHASTIC PROCESSES AND SAMPLING

6.0. Introduction. A stochastic process, in general, can be defined as a collection of stochastic variables. In the last two chapters we defined *s. v*'s and studied some special probability distributions. The study of a collection of stochastic variables having special properties is very useful in many branches of applied statistical mathematics. Most of the applications of this particular study of a collection of *s.v*'s is in the description and analysis of the development of a random quantity over time. Hence the name 'as stochastic process' is often identified with a collection of *s. v*'s $X(t)$ where $t \in T$ and T is a set of real numbers. In our discussion we will not be interested in a process over time, but we will simply consider a collection of *s. v*'s having some common properties. For a study of *s.v*'s over time the reader is advised to see any book on Stochastic Processes.

6·1. SUMS OF STOCHASTIC VARIABLES

In the statistical theory of distribution and testing hypotheses sums and linear combinations of *s.v*'s play an important role. So we will investigate some of the properties of sum of *s.v*'s before defining special collection of *s.v*'s such as a simple random sample etc.

Theorem 6.1. If $X_1, X_2 \ldots, X_k$ are k independent *s.v*'s with M.G.F's. $M_{X_i}(t)$ for $i=1, 2, \ldots, k$ and if $Y=X_1+\ldots+X_k$ then the M.G.F. of Y is the product of the M.G.F's of $X_1, X_2, \ldots, X_k$.

i.e.,

$$M_Y(t)=M_{X_1}(t)\, M_{X_2}(t) \ldots M_{X_k}(t)$$

$$= \prod_{i=1}^{k} M_{X_i}(t) \tag{6·1}$$

where Π is a notation for products. For example

$$\prod_{i=1}^{n} a_i = a_1 a_2 \ldots\ldots a_n$$

Proof. $M_Y(t)=Ee^{tY}=E\, e^{t(X_1+X_2+\ldots X_k)}$

$$=E(e^{tX_1} \,.\, e^{tX_2} \cdots e^{tX_k})$$

$$=E\, e^{tX_1} \,.\, E\, e^{tX_2} \ldots E\, e^{tX_k}$$

(by applying theorem 5.1 repeatedly)

$$=M_{X_1}(t) \cdot M_{X_2}(t) \ldots M_{X_k}(t)$$

$$=\prod_{i=1}^{k} M_{X_i}(t).$$

Ex. 6.1.1. *Examine the distribution of* $Y=X_1+X_2+\ldots+X_n$ *when (a)* X_i *has a Binomial distribution with parameters* N_i *and* p *for* $i=1, 2, \ldots\ldots k$, *(b)* X_i *has a Poisson distribution with parameter* λ_i *for* $i=1, 2, \ldots\ldots k$, *(c)* X_i *has a Gamma distribution with parameters* α_i *and* β *for* $i=1, 2, \ldots\ldots k$, *(d)* X_i *has a normal distribution with parameters* μ_i *and* σ_i *for* $i=1, 2, \ldots k$, *where* $X_1, X_2, \ldots\ldots, X_k$ *are independent.*

Sol. (*a*) $\quad M_{X_i}(t)=(q+p\,e^t)^{N_i},\ i=1, 2, \ldots k$

Since the X's are independent, by Theorem 5.2

$$M_Y(t)=(q+p\,e^t)^{N_1} \cdot (q+p\,e^t)^{N_2} \ldots (q+p\,e^t)^{N_k}$$

$$=(q+p\,e^t)^{N_1+N_2+\ldots+N_k} \qquad (6.2)$$

By the uniqueness property of M.G.F. Y has a Binomial distribution with parameters $N=(N_1+N_2+\ldots+N_k)$ and p.

(*b*) $\quad M_{X_i}(t)=e^{\lambda_i(e^t-1)}$

$$\therefore \quad M_Y(t)=\prod_{i=1}^{k} M_{X_i}(t)=e^{(\lambda_1+\lambda_2+\ldots+\lambda_k)(e^t-1)}$$

(by theorem 6.1) $\qquad$ (6.3)

$\therefore$ Y has a Poisson distribution with the parameter $(\lambda_1+\ldots+\lambda_k)$.

(*c*) $\quad M_{X_i}(t)=(1-\beta t)^{-\alpha_i}$

$$\therefore \quad M_Y(t)=(1-\beta t)^{-\alpha_1-\alpha_2-\ldots-\alpha_k} \quad \text{(by theorem 5.2)} \qquad (6{\cdot}4)$$

$\therefore$ Y has a Gamma distribution with parameters $(\alpha_1+\alpha_2+\ldots+\alpha_k)$ and β.

(*d*) $\quad M_{X_i}(t)=e^{t\mu_i+t^2\sigma_i^2/2}$

$$\therefore \quad M_Y(t)=\prod_{i=1}^{k} e^{t\mu_i+t^2\sigma_i^2/2}$$

$$e^{t(\mu_1+\ldots+\mu_k)+(\sigma_1^2+\sigma_2^2+\ldots+\sigma_k^2)\frac{t^2}{2}} \qquad (6{\cdot}5)$$

$\therefore$ Y has a normal distribution with mean $(\mu_1+\mu_2+\ldots+\mu_k)$ and with standard deviation equal to square root of $\Big(\sigma_1^2+\sigma_2^2+\ldots$

σ_k^2), or with variance $\sigma_1^2+\sigma_2^2+\ldots+\sigma_k^2$. For any *s.v*'s $X_1, X_2 \ldots X_k$ the probability function of $Y=X_1+X_2+\ldots+X_k$ is also called the convolution of $X_1, X_2, \ldots X_k$.

6.2. LINEAR COMBINATION OF STOCHASTIC VARIABLES

Let $$Y=a_1X_1+a_2X_2+\ldots+a_kX_k$$

where $a_1, a_2, \ldots\ldots a_k$ are constants. Y is called a linear combination of the stochastic variable $X_1, X_2, \ldots X_k$.

Let $E(X_i)=\mu_i$ and $Var(X_i)=\sigma_i^2$ for $i=1, 2, \ldots k$.

$$E(Y)=E[a_1X_1+a_2X_2+\ldots+a_kX_k]$$
$$=a_1E(X_1)+a_2E(X_2)+\ldots+a_kE(X_k)$$

(the proof of this step is left to the reader)

$$=a_1\mu_1+a_2\mu_2+\ldots\ldots+a_k\mu_k$$
$$=\sum_{i=1}^{k} a_i \mu_i \tag{6.7}$$

$$Var(Y)=E[Y-E(Y)]^2$$
$$=E[a_1X_1+\ldots+a_kX_k-a_1\mu_1-\ldots-a_k\mu_k]^2$$
$$=E[a_1(X_1-\mu_1)+a_2(X_2-\mu_2)+\ldots+a_k(X_k-\mu_k)]^2$$
$$=E[\sum_{i=1}^{k} a_i^2 (X_i-\mu_i)^2+\sum_{\substack{i, j\\ i\neq j}} a_i a_j(X_i-\mu_i)(X_j-\mu_j)]$$
$$=\sum_{i=1}^{k} a_i^2 E(X_i-\mu_i)^2+\sum_{i\neq j} a_ia_j . Cov. (X_i, X_j)$$
$$=\sum_{i=1}^{k} a_i^2 Var(X_i)+\sum_{i\neq j} a_ia_j Cov(X_i, X_j) \tag{6.8}$$

This may also be written as

$$Var(Y)=\sum_{i=1}^{k} a_i^2 Var(X_i)+2\sum_{i<j} a_ia_j Cov(X_i, X_j) \tag{6·9}$$

Note. When the X's are independent the various covariances are zeros and hence

$$Var(Y)=\sum_{i=1}^{k} a_i^2 Var(X_i). \tag{6·10}$$

$\sum_{i \neq j} a_i a_j$ Cov (X_i, X_j) means that the sum of all the terms where $i \neq j$, or except terms of the form

$$a_1 a_1 \text{ cov } (X_1, X_1),\ a_2\, a_2 \text{ cov } (X_2, X_2), \text{ etc.}$$

$$\text{Cov } (X_i, X_j) = \rho_{ij}\ \sigma_i \sigma_j$$

(where ρ_{ij} denotes the correlation between X_i and X_j)

Var (Y) may also be written as

$$\text{Var}(Y) = \sum_{i=1}^{k} a_i^2\ \sigma_i^2 + \sum_{i \neq j} a_i a_j\ \rho_{ij}\ \sigma_i \sigma_j.$$

Ex. 6.2.1. *Evaluate the expected value and standard deviation of $\overline{X} = (X_1 + X_2 + \ldots\ldots + X_n)/n$ where $X_1, X_2, \ldots X_n$ are independent and $E(X_i) = \mu$ and $Var\ (X_i) = \sigma^2$ for $i = 1, 2, \ldots\ n$ ($\overline{X}$ is read as X bar).*

Sol.
$$E(\overline{X}) = [(X_1 + X_2 + \ldots + X_n)/n]$$
$$= \frac{1}{n}[E(X_1) + E(X_2) + \ldots + E(X_n)]$$
$$= \frac{1}{n}[\mu + \mu + \ldots + \mu] = n\mu/n = \mu \qquad (6 \cdot 11)$$
$$\text{Var } (\overline{X}) = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 \text{Var } (X_i) + 0$$
$$= \frac{1}{n^2}[\sigma^2 + \sigma^2 + \ldots + \sigma^2]$$
$$= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$\therefore$ The standard deviation of $\overline{X} = \sigma/\sqrt{n}$. $\qquad (6 \cdot 12)$

Exercises

6.1. If $X_1\ X_2 \ldots\ldots X_n$ are independently and identically distributed point Binomial variates show that $Y = X_1 + \ldots + X_n$ is a Binomial variate. If X is a *s.v.* taking only values 1 and 0 with probabilities p and $1-p$ respectively then X is called a point Binomial variate. Show that

$$\frac{X - np}{\sqrt{np(1-p)}} \rightarrow N(0, 1)$$

where $n \rightarrow \infty$.

[**Hint.** Use the central limit theorem.]

6.2. For a point Binomial variate show that the probability of getting exactly x successes is given by

$$f(x) = \begin{cases} p^x (1-p)^{1-x} \text{ for } x = 0, 1 \\ 0 \text{ elsewhere.} \end{cases}$$

6.3. Three independent *s.v*'s X_1, X_2 and X_3 have (1) binomial distributions with parameters $p_1=p_2=p_3=1/2$, $N_1=10$, $N_2=15$, $N_3=30$; (2) Poisson distributions with parameters $\lambda_1=\lambda_2=3$ and $\lambda_3=5$, obtain the distribution of $X_1+X_2+X_3$.

6.4. If two independent *s. v*'s X_1 and X_2 have the same rectangular distributions, obtain the distribution of $(X_1+X_2)/2$.

6.5. If n independent *s.v*'s have gamma distributions with the parameters $\beta_1=\beta_2=\ldots=\beta_n=3$, $\alpha_1=a$, $\alpha_2=2a,\ldots$, $\alpha_n=na$, obtain the distribution of the sum of the *s.v*'s.

6.6. If X_1, X_2, X_3 are independently and normally distributed as $N(\mu_1, \sigma_1)$, $N(\mu_2, \sigma_2)$ and $N(\mu_3, \sigma_3)$ respectively, obtain the distribution of (*a*) $X_1-2X_2+X_3$, (*b*) $X_1+X_2-5X_3$.

6.7. If X is the number of successes in N independent trials, where the probability of success in the *i*th trial is p_i, obtain E(X) and Var (X).

6.3. SAMPLES FROM THEORETICAL POPULATIONS

In chapter 1 we defined a statistical population as a set where every element may be characterized by one or more characteristics. If the population is univariate then every element is characterized by one characteristic. For example the height measurements of all the students in a particular country at a particular time. This is an example of a univariate finite population. Let us consider a set where the elements are defined by a probability function $f(x)$ or by a stochastic variable X, in the sense that the probability of getting a particular element x_0 of the set is given by $f(x_0)$ if X is a discrete variable and the probability of getting an element in the neighbourhood of x_0 is given by

$$\int_{x_0-\frac{\Delta x_0}{2}}^{x_0+\frac{\Delta x_0}{2}} f(x)dx=f(x_0)d\alpha_0$$

approximately, where Δx_0 denotes a small neighbourhood at x_0 if X is a continuous stochastic variable. If a set is thus defined by a stochastic variable X or by a probability law $f(x)$ we say that such a set is a theoretical population designated by the stochastic variable X. This is a univariate case. The same ideas may be generalized for defining a multivariate theoretical population. For example a normal population means a set designated by a normal variable X or by the normal probability law

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\ \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

where μ and σ are parameters. If x_0 is called an observation from a population $f(x)$, this means that x_0 is an element of the set designated by the stochastic variable X having the probability

function $f(x)$. Here x_0 may be considered to be the value assumed by a stochastic variable X having a probability function $f(x)$. If x_1 and x_2 are two independent observations from $f(x)$ this may be considered to be the values assumed by two independent stochastic variables X_1 and X_2, each having the same distribution given by the probability function $f(x)$.

Now we will define a simple random sample from a theoretical population designated by a probability function $f(x)$. A set of stochastic variables $X_1, X_2, \ldots X_n$ which are independently and identically distributed as $f(x)$, is called a simple random sample of size n from the population $f(x)$. For example if $X_1, X_2, \ldots X_n$ are independently and identically distributed as a normal distribution with parameters μ and σ then $X_1, X_2, \ldots, X_n$ is called a simple random sample from a $N(\mu, \sigma)$. If $X_1 \ldots X_n$ is a random sample from $f(x)$ then their joint probability function,

$$f(x_1, x_2, \ldots x_n)=f(x_1) \cdot f(x_2) \ldots f(x_n).$$

[Here $f(x_i)$ means $f(x)$ at $x=x_i$, for $i=1, 2, \ldots n$).

$$=(2\pi\sigma^2)^{-n/2} e^{-\sum_{i=1}^{n} (x_i-\mu)^2/2\sigma^2}$$

if $$f(x)=(2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/2\sigma^2}.$$

(Since $X_1, \ldots, X_n$ are independently and identically distributed).

6.31. Statistics. If $X_1, \ldots, X_n$ is a simple random sample from a population $f(x)$ [*i.e.*, $X_1, \ldots, X_n$ are independently and identically distributed as $f(X)$] then any single valued function of $X_1, \ldots, X_n$ is called Statistic. (Plural of statistic is statistics and is different from the science Statistics or a collection of observations.) This definition can be extended for a multivariate case also. For example,

$$\overline{X} =(X_1+\ldots+X_n)/n \text{ is statistic ;}$$

$$S^2=\sum_{i=1}^{n} (X_i-\overline{X})^2/n \text{ is a statistic etc.}$$

An infinite number of statistics can be constructed from a given sample. Evidently a statistic has its own distribution if it is a stochastic variable. Any function $\phi(X)$ of a *s.v.* X need not be a *s.v.* But in the following discussions we will consider only Borel measurable functions of a *s.v.*, which are *s.v.*'s. Further from the definition of a statistic any X_i itself is a statistic.

The distribution of a statistic is called a sampling distribution. For example the distribution of $\overline{X}$ is a sampling distribution. The distribution of the largest of $X_1, \ldots, X_n$ is a sampling distribution etc. In estimation theory these statistics are also called

estimators. Estimators and their properties will be discussed in the chapter on estimation.

If we have a sample from a multi-variate population we can construct a statistic for this sample in a similar way as in a uni-variate case. For example let $(X_1, Y_1), (X_2, Y_2),\ldots(X_n, Y_n)$ be a random sample from a bivariate population [this means that the random variables (X_i, Y_i) have a joint distribution $f(x, y)$ for $i=1, 2,\ldots n$ and $(X_1, Y_1), (X_2, Y_2)\ldots(X_n, Y_n)$ are all independently and identically distributed as $f(x, y)$] then the sample covariance

$$S_{12}=\sum_{i=1}^{n}(X_1-\bar{X})(Y_i-\bar{Y})/n \text{ is a statistic.}$$

The sample correlation coefficient

$$r_{12}=\frac{S_{12}}{S_1.S_2} \text{ is a statistic}$$

where $S_1^2=\sum\limits_{i=1}^{n}(X_1-\bar{X})^2/n$ and $S_2^2=\sum\limits_{i=1}^{n}(Y_i-\bar{Y})^2/n$ are the sample variances. A number of statistics may be constructed from a sample from a multi-variate population.

6·32. The Sample Mean. If $X_1, X_2,\ldots X_n$ is a random sample from a population $f(x)$ then

$$\bar{X}=(X_1+\ldots+X_n)/n$$

is called the sample mean. Sample mean is evidently a Statistic. Let us evaluate the expected value and variance of the sample mean $\bar{X}$

$$E(\bar{X})=\mu \tag{6·13}$$

and Var $(\bar{X})=\sigma^2/n$ (See Ex. 5·6·1.) (6·14)

Therefore, $\bar{X}$ has the mean μ and the standard deviation $\sigma/\sqrt{n}$ where n is the sample size. For example if a random sample of size n is taken from an exponential distribution with parameter θ then the sample mean has the expected value θ and a variance θ^2/n since the mean and variance of a stochastic variable having an exponential distribution, are θ and θ^2 respectively.

6·33. The Sample Variance. If $X_1, X_2,\ldots X_n$ is a random sample from a population with mean μ and variance σ^2 then the sample variance S^2 is defined as

$$S^2=\sum_{i=1}^{n}(X_i-\bar{X})^2/n=\sum_{i=1}^{n}\frac{X_i^2}{n}-(\bar{X})^2 \tag{6·15}$$

where $\bar{X}$ is the sample mean. Let us evaluate the expected value of the sample variance

$$E(S^2)=E\sum_{i=1}^{n}(X_i-\bar{X})^2/n=E\sum_{i=1}^{n}(X_i-\mu+\mu-\bar{X})^2/n$$

$$=E\left[\sum_{i=1}^{n}(X_i-\mu)^2+n(\bar{X}-\mu)^2-2\Sigma_i(\bar{X}_i-\mu)(X-\mu)\right]/n$$

$$=\left[\sum_{i=1}^{n}E(X_i-\mu)^2-n\,E(\bar{X}-\mu)^2\right]/n \tag{6·16}$$

But $E(X_i)=\mu$, $E(\bar{X})=\mu$ and Var $(\bar{X})=\sigma^2/n$

$$\therefore \quad E(S^2)=\frac{1}{n}\,[n\,\sigma^2-n.\sigma^2/n]=\frac{(n-1)}{n}\sigma^2$$

$$\therefore \quad E\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{n}=\frac{(n-1)}{n}\,\sigma^2=E(S^2) \tag{6·17}$$

$$\therefore \quad E\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{n-1}=\sigma^2=E\left[\frac{n}{(n-1)}\,S^2\right]. \tag{6·18}$$

Here we say that $nS^2/(n-1)$ is an unbiased statistic for σ^2, in the sense that $E[nS^2/(n-1)]$ is σ^2. This aspect of unbiasedness of statistics will be discussed in the chapter on point estimation. Because of this property some authors define the sample variance as $\sum_{i=1}^{n}(X_i-\bar{X})^2/(n-1)$. But we will follow the definition in section 6·33.

Ex. 6.33.1. *Given that 5, 8, 4, 3 is an observed random sample of size 4 from a population $f(x)$, evaluate the sample mean and the sample variance.*

Sol. These observations 5, 8, 4, 3 are from the population $f(x)$. This means that 5, 8, 4, 3 are the values assumed by the stochastic variables X_1, X_2, X_3, and X_4 respectively, where X_1, X_2, X_3 and X_4 are independently and identically distributed as $f(x)$.

The sample mean $\bar{X}=(X_1+X_2+\ldots+X_n)/n$.

Therefore an observed value of this stochastic variable $\bar{X}$ is given by $(5+8+4+3)/4=5$.

For this observed sample the sample mean is 4 similarly the sample variance for the given sample is

$$=[(5-5)^2+(8-5)^2+(4-5)^2+(3-5)^2]/4$$
$$=14/4.$$

This may be taken as a value assumed by the stochastic variable S^2.

6·34. The Standard Error. From a random sample, we can construct an infinite number of statistics. The standard deviation (square root of the variance) of a statistic is known as the standard error of the statistic. For example in Section 6·32 we have seen that the statistic $\overline{X}$, (sample mean), has a standard deviation $\sigma/\sqrt{n}$. Hence the standard error of the sample mean is $\sigma/\sqrt{n}$ where σ is the population standard deviation and n is the sample size.

Ex. 6·34·1. *Show that $\overline{X}-\overline{Y}$ has a standard error*

$$\left(\sigma_1^2/n_1+\sigma_2^2/n^2\right)^{1/2}$$

where $\overline{X}$ is the sample mean of a sample of size n_1 from a population with mean μ_1 and variance σ_1^2 and $\overline{Y}$ is the sample mean of a sample of size n_2 from a population with mean μ_2 and variance σ_2^2. where the populations are assumed to be independent.

Sol. Let $T=\overline{X}-\overline{Y}$

$$E(T)=E(\overline{X}-\overline{Y})=E(\overline{X})-E(\overline{Y})$$
$$=\mu_1-\mu_2 \tag{6·19}$$
$$\text{Var. }(T)=E[T-E(T)]^2$$
$$=E[\overline{X}-\overline{Y}-(\mu_1-\mu_2)]^2$$
$$=E[(\overline{X}-\mu_1)-(\overline{Y}-\mu_2)]^2$$

$=E(\overline{X}-\mu_1)^2+E(\overline{Y}-\mu_2)^2$ (the covariance term vanishes since the populations are given to be independent)

$$=\sigma_1^2/n_1+\sigma_2^2/n_2 \tag{6·20}$$

$\therefore$ The standard deviation of T is $\left(\sigma_1^2/n_1+\sigma_2^2/n_2\right)^{1/2}$

Comments. When $\mu_1=\mu_2$ and $\sigma_1=\sigma_2$, *i.e.*, if we take two independent samples of sizes n_1 and n_2 from the same population then $\overline{X}-\overline{Y}$ has the standard error

$$(\sigma^2/n_1+\sigma^2/n_2)^{1/2}=\sigma(1/n_1+1/n_2)^{1/2} \tag{6·21}$$

If two stochastic variables X and Y are independent, we say that the populations represented by X and Y are independent (statements like two independent exponential populations with parameters θ_1 and θ_2, two independent normal populations with parameters μ_1 and σ_1 and μ_2 and σ_2 etc.).

Ex. 6·34·2. *If $\bar{X}$ and $\bar{Y}$ denote the proportion of successes (total number of successes divided by the total number of trials) in two independent Binomial probability situations with parameters p_1 and N_1 and p_2 and N_2 respectively, find the expected value and the standard error of the statistic $T=\bar{X}-\bar{Y}$.*

Sol. If X denotes the number of successes in a Binomial situation with parameters p and N then the proportion of success $Z=X/N$ has an expected value p and a variance $p(1-p)/N$. This is easily seen, since $Z=X/N=(1/N).X$ and $1/N$ is a constant. Hence $E(Z)=(1/N).E(X)=(1/N).Np=p$ and $\text{Var}(Z)=(1/N)^2\ \text{Var}(X)=(1/N)^2\ N.p(1-p)=p(1-p)/N$

$$T=\bar{X}-\bar{Y}$$

$$\therefore \quad E(T)=E(\bar{X})-E(\bar{Y})=p_1-p_2 \tag{6·22}$$

$\text{Var}(T)=\text{Var}(\bar{X})+\text{Var}(\bar{Y})$ (Since the Binomial populations are given to be independent).

$$=p_1(1-p_1)/N_1+p_2(1-p_2)/N_2 \tag{6·23}$$

$\therefore$ The standard error of T is

$$[p_1(1-p_1)/N_1+p_2(1-p_2)/N_2]^{1/2}$$

In example 6·1·1 we have seen that if $X_1, X_2, \ldots X_n$ are independent and if $X_i : N(\mu_i, \sigma_i)$ for $i=1, 2, \ldots n$ then $Y=X_1+X_2+\ldots X_n$ has a normal distribution with parameters, mean$=\mu_1+\mu_2+\ldots+\mu_n$ and variance$=\sigma_1^2+\sigma_2^2+\ldots+\sigma_n^2$. As a special case of this, if $X_1, X_2, \ldots X_n$ is a random sample from $N(\mu, \sigma)$ then

$$\bar{X}=(X_1+X_2+\ldots+X_n)/n$$

has a normal distribution with mean μ and with variance σ^2/n *i.e.*, $\bar{X} : N(\mu, \sigma/\sqrt{n})$ if $X_1, \ldots X_n$ is a random sample from $N(\mu, \sigma)$.

Hence

$$Z=\frac{\bar{X}-E(\bar{X})}{\sqrt{\text{Var}(\bar{X})}}=\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}$$

has a normal distribution with zero mean and unit variance *i.e.*,

$$Z : N(0, 1)$$

This means that if a random sample of size n is taken from $N(\mu, \sigma)$ then the standardized sample mean has a standardized

normal distribution, whatever the sample size n may be. Now we will examine the distribution of the standardized sample mean if a random sample of size n is taken from any population $f(x)$. One form of the central limit theorem gives the distribution of the standardized sample mean for any parent distribution.

6.35. The Central Limit Theorem.

Theorem 6·2. If $X_1, X_2, \ldots X_n$ is a random sample from a population $f(x)$ with finite mean and variance then

$$Z = [\bar{X} - E(\bar{X})] \sqrt{\text{Var}\,(\bar{X})}$$

has a distribution which approximates to a standardized normal distribution when the sample size is sufficiently large

i.e., If $$Z = \frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}$$

then $$Z : N(0, 1) \text{ when } n \to \infty \tag{6·24}$$

where $\mu = E(X_i)$ and $\sigma^2 = \text{Var}\,(X_i)$ for $i = 1, 2, \ldots n$.

Proof. Let $M_X(t)$ be the M.G.F. of the parent population.

or $$M_{X_i}(t) = M_X(t) \text{ for } i = 1, 2, \ldots n$$

Let $$Y = (X_1 + X_2 + \ldots + X_n), \text{ then}$$

$$M_Y(t) = [M_X(t)]^n \text{ (Theorem 6.1)} \tag{6 25}$$

$$\therefore \quad M_Z(t) = e^{-\frac{\mu\sqrt{n}\,t}{\sigma}} \left[M_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

$$\therefore \quad \log M_Z(t) = -\mu\sqrt{n}\,t/\sigma + n \log M_X\left(\frac{t}{\sigma\sqrt{n}}\right) \tag{6·26}$$

(where 'log' denotes the natural logarithm).

But $$M_X\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \mu'_1\left(\frac{t}{\sigma\sqrt{n}}\right) + \frac{\mu'_2}{2!}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \ldots\ldots$$

where $\mu_1' = \mu$, $\mu'_2 \ldots$ are the raw moments of X

$$= 1 + T \text{ (say)} \tag{6·27}$$

$$\log(1+T) = T - T^2/2 + T^3/3 - \ldots$$

(assuming the validity of the expansion) (6·28)

$$\therefore \quad \log M_Z(t) = -\mu\sqrt{n}\,t/\sigma + n\left[T - \frac{T^2}{2} + \frac{T^3}{3} + \ldots\right]$$

where $$T = \mu_1'(t/\sigma\sqrt{n}) + \frac{\mu_2'}{2!}(t/\sigma\sqrt{n})^2 + \ldots \tag{6·29}$$

Now collecting the coefficients of $t, t^2, \ldots$ in $\log M_Z(t)$ we get

$$\log M_Z(t) = (-\mu\sqrt{n} + \mu_1'\sqrt{n})(t/\sigma) +$$

$$+\left(\frac{\mu'_2}{2}-\frac{(\mu_1')^2}{2}\right)\left(\frac{t}{\sigma}\right)^2+\left(\frac{\mu'_3}{6\sqrt{n}}-\frac{\mu'_1\mu'_2}{2\sqrt{n}}+\frac{(\mu_3')^3}{3\sqrt{n}}\right)(t/\sigma)^3+\ldots \quad (6.30)$$

Since $\mu'_1=\mu$ the coefficient of t/σ is zero, and the coefficients of $(t/\sigma)^3$, $(t/\sigma)^4$,... contain powers of n in the denominators. For example the coefficient of $(t/\sigma)^3$ contain $\sqrt{n}$ in the denominator etc. Hence, when $n\to\infty$.

$$\log M_Z(t)\to\tfrac{1}{2}(\mu'_2-\mu_1'^2)(t/\sigma)^2=\frac{\mu_2}{2\sigma^2}t^2=t^2/2 \quad (6\cdot31)$$

i.e. $M_Z=e^{t^2/2}$

when $n\to\infty$. (6·32)

By the uniqueness property of moment generating functions, Z is N(0, 1) when $n\to\infty$ since N(0, 1) has a M.G.F.$=e^{t^2/2}$. This completes the proof.

The reader might have already realized the importance of this theorem. According to the theorem if $X_1, X_2,\ldots X_n$ is a random sample from an exponential population with parameter θ, (sec. 4·12), then

$$Z=\frac{\bar{X}-E(\bar{X})}{\sqrt{\text{Var}(\bar{X})}}=\frac{(\bar{X}-\theta)}{\theta/\sqrt{n}}$$

is approximately N(0, 1) when n is sufficiently large. A good approximation when the parent population is symmetric is usually obtained when $n\geqslant 30$. For a detailed discussion of the various central limit theorems and other theorems on convergence of stochastic variables see M. Loeve, Probability Theory, D. Van Nostrand Co, Inc, New York, 1955.

Exercises

6.8. Obtain the distribution of $\bar{X}$ when the sample of size n is from a N(μ, σ) and n has a binomial distribution with the parameters p and N.

6.9. If $X_1, \ldots X_n$, is a random sample from a N(μ, σ) and n has a binomial distribution with the parameters p and N obtain (1) E (Y), (2) Var (Y) where $Y=X_1+\cdots+X_n$. (see problem 5.15).

6.10. If (X_1, X_2, X_3) is a random sample from a N(μ, σ) obtain the distribution of $X_1-X_2+X_3$.

6.11. If (2, 3), (4, 7), (1, −1), (0, −5), (5, 8) is an observed random sample from a bivariate population, obtain the sample means, sample variances and the sample correlation coefficient.

6.12. If (X_1, Y_1), (X_2, Y_2) is a random sample of size 2 from a bivariate population obtain the following statistics ; (1) the sample means, (2) the sample variances, (3) the sample covariance.

6.13. If three random samples of sizes $n_1, n_2\ n_3$ are taken from three independent populations with means μ_1, μ_2, μ_3 and with variances

σ_1^2, σ_2^2 and σ_3^2 respectively, what is the standard error of the statistics $X-Y+2Z$ where X, Y and Z denote the sample means.

6.14. If two independent samples of sizes n_1 and n_2 are taken from the same population with mean μ and variance σ^2, what is the standard error of the statistic $X-2Y$, where X and Y denote the sample means.

6.15. Two random samples of sizes 10 and 20 are taken from a population with variance $\sigma^2=16$. By using Chebyshev's theorem obtain the limit for the probability that the sample means will not differ by more than 3 units.

6.16. Two independent random samples of sizes 20 and 30 are taken from two populations with the parameters $\mu_1=\mu_2$ and $\sigma_1=10$ and $\sigma_2=4$. Using Chebyshev's theorem what can we assert with a probability of at least 0.80, about the possible difference of the sample means.

6.17. Ten per cent of the oranges in a grocery store are spoiled. A housewife picks up 20 oranges at random. Obtain an upper bound for the probability that the proportion of spoiled oranges in this sample differs from the true proportion by more than 1%.

6.18. A random sample of size 50, taken from a population with parameters $\mu=20$ and $\sigma=4$, has a sample mean 25. If random samples of size 50 are taken, what is the probability of getting a sample mean as large as 25.

6.19. Two independent random samples of size 100 each are taken from two populations with the parameters $\mu_1=20$, $\mu_2=20$, $\sigma_1=5$, $\sigma_2=6$. If sampling is continued, what is the probability of getting two sample means which differ at the most by 5.

6.4. ORDER STATISTICS

We will develop the theory for the case when the parent population is continuous. The ideas may be extended to a discrete population as well. Let $x_1, x_2, \ldots x_n$ be an observed random sample from a continuous population $f(x)$. These observations may be ordered according to the order of their magnitudes. Let $u_1 \leqslant u_2 \leqslant \ldots \leqslant u_n$ be the arrangement of $x_1, x_2, \ldots x_n$; where u_1 is the smallest one, u_n is the largest one and u_r is the r^{th} largest one. $u_1, u_2, \ldots, u_n$ may be considered to be the values assumed by the stochastic variables $U_1, U_2, \ldots U_n$. For example if we take a number of random samples of size n and order the observations according to their magnitudes in each sample, we get a number of values assumed by $U_1, U_2, \ldots U_n$. The stochastic variable U_r is called the r^{th} order statistic and the distribution of U_r is called the sampling distribution of the r^{th} order statistic. u_r is the r^{th} largest of the x's and hence $(r-1)$ of the x's fall below u_r and $n-r$ of the x's fall above u_r. For convenience we assume that there is only one observation which is the r^{th} largest.

Let p_1, p_2, and p_3 be the probabilities that an observed value of the stochastic variable X falls below u_r, in the interval (u_r, u_r+h), and in the interval (u_r+h, ∞) respectively, where h is

a very small quantity or (u_r, u_r+h) is a very small interval at u_r (see Fig. 6.1).

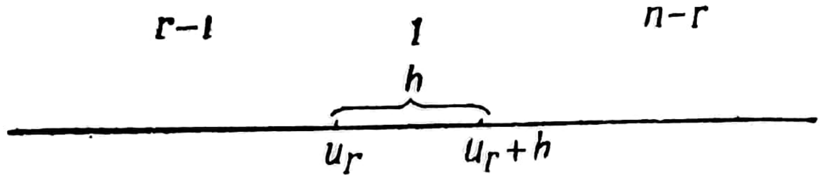


Fig. 6.1.

$$p_1=\int_{-\infty}^{u_r} f(x)dx=F(u_r) \tag{6.33}$$

$$p_2=\int_{u_r}^{u_r+h} f(x)dx=f(\eta)\,.\,h$$

where $\qquad u_r \leqslant \eta \leqslant u_r+h, \qquad$ (6.34)

(by the mean value theorem)

$$p_3=\int_{u_r+h}^{\infty} f(x)dx=1-F(u_r+h) \tag{6.35}$$

when $h \to 0$, $p_1=F(u_r)$, $p_2=f(u_r)$ where $f(u_r)$ is the density function $f(x)$ at $x=u_r$ (not the density function of U_r) and $p_3=1-F(u_r)$. We have observed $r-1$ of the x's below u_r, one x at u_r and $n-r$ of the x's above u_r. Hence by the multinomial distribution the density function of U_r is

$$g(u_r)=\frac{n\,!}{(r-1)\,!\;1\,!\;(n-r)\,!}\;p_1^{r-1}\;p_2^{1}\;p_3^{n-r}$$

$$=\frac{n\,!}{(r-1)\,!\;(n-r)\,!}\left[\int_{-\infty}^{u_r} f(x)dx\right]^{r-1} f(u_r).$$

$$\left[\int_{u_r}^{\infty} f(x)dx\right]^{n-r} \tag{6·36}$$

where $g(u_r)$ is the density function of U_r and $f(u_r)$ is $f(x)$ at $x=u_r$. For example the density functions of the smallest order statistic U_1 and the largest order statistic U_n are given as

$$g(u_1)=g(u_r) \quad \text{for } r=1,=\frac{n\,!}{(n-1)\,!}f(u_1)\left[\int_{u_1}^{\infty} f(x)dx\right]^{n-1}$$

$$=n.\ f(u_1)\left[\int_{u_1}^{\infty} f(x)dx\right]^{n-1} \text{ for } -\infty<u_1<\infty. \tag{6·37}$$

$$g(u_n)=n.\ f(u_n)\ .\left[\int_{-\infty}^{u_n} f(x)dx\right]^{n-1} \text{ for } -\infty<u_n<\infty. \tag{6·38}$$

6.41. The joint distribution of U_r and U_s where $r\neq s$. The joint distribution of two ordered statistics U_r and U_s where $r\neq s$, may be obtained in a similar fashion. This is illustrated in Fig. 6.2.

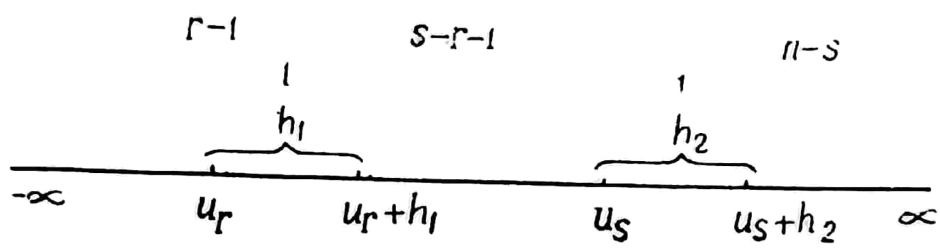


Fig. 6·2.

The joint density function of U_r and U_s is

$$g(u_r, u_s)=\frac{n\,!}{(r-1)\,!\ (s-r-1)\,!\ (n-s)\,!}\left[\int_{-\infty}^{u_r} f(x)dx\right]^{r-1}$$

$$f(u_r)\left[\int_{u_r}^{u_s} f(x)dx\right]^{s-r-1} f(u_s)\ .\left[\int_{u_s}^{\infty} f(x)dx\right]^{n-s}$$

$$\text{for } -\infty<u_r<u_s<\infty. \tag{6·39}$$

The sampling distributions of the median and the range are obtained from equations (6·36) and (6·39) respectively. For example if we have $2m+1$ observations the $(m+1)^{th}$ ordered observation is the median of the observations and hence the distribution of the median M is obtained by replacing n by $2\,m+1$ and r by $m+1$ in equation (6·36). U_n-U_1 is the sample range and hence the distribution of the range R may be obtained from the joint distribution of U_n and U_1. Sometimes it may be difficult to obtain the exact values of the integrals in the equations. In such cases we may obtain approximate values by some approximation procedures.

Ex. 6.4.1. *Obtain the distribution of the smallest order statistic U_1 from an exponential population with the parameter θ.*

Sol.
$$f(x)=\frac{1}{\theta}\ e^{-x/\theta} \text{ for } 0<x<\infty.$$
$$=0 \text{ elsewhere}$$

$$\therefore \quad \int_{u_1}^{\infty} f(x)dx = e^{-u_1/\theta} \tag{6·40}$$

$\therefore$ The density function for the smallest order statistic U_1 is

$$g(u_1) = n \,.\, f(u_1)\left[\int_{u_1}^{\infty} f(x)\,dx\right]^{n-1}$$

$$= n \,.\, \frac{1}{\theta}\; e^{-u_1/\theta}\,[e^{-u_1/\theta}]^{n-1}$$

$$= \frac{n}{\theta}\; e^{-(n/\theta)u_1} \text{ for } 0 < u_1 < \infty$$

$$= 0 \text{ elsewhere.} \tag{6·41}$$

Comments. The distribution of U_1 is again an exponential distribution with the parameter $1/\theta$ multiplied by the sample size n. This is a characteristic property of some distributions. A family or a set of distributions may be characterized by using this property.

Exercises

6.20. Obtain the distribution of the smallest and the largest order statistics from a sample of size n if the parent population is,

$$f(x, \theta) = \begin{cases} 1/(\beta-\alpha) & \text{for } \alpha < x < \beta,\ \alpha > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

6.21. For the distribution in problem 6·20 obtain the distribution of the sample range $(U_n - U_1)$, by obtaining the joint distribution of U_r and U_s.

6.22. Obtain the distribution of the sample median if a random sample of size $2n+1$ is taken from a normal population.

6.23. For large n it can be proved that the distribution of the sampel median m of a sample of size $2n+1$ is approximately normal with mean the population median M and with variance $1/8n[f(\text{M})]^2$ where $f(\text{M})$ is the population density function $f(x)$ at $x=\text{M}$. Obtain the variance of m if $f(x)$ is a $N(\mu, \sigma)$, and if the sample size is large.

6.24. Obtain the distribution of the sample range R if samples of size n are taken from an exponential population. Compare the distribution with the parent distribution when $n=2$.

6.5. SAMPLING FROM A FINITE POPULATION

The following is a note on sampling from a finite population For a detailed discussion see any book on Sampling. Definitions and examples of univariate as well as multivariate populations. are given in chapter 1 and in section 6.3. Let us consider a finite univariate population given by a set of observations or data. For

example, the annual incomes of all the citizens in a city in a particular year, the height measurements of all the students in the universities in a particular country at a particular time, the set of bullets produced by a machine in a certain time interval etc., all form univariate finite populations. If we are interested in studying some population characteristics such as the mean μ, a measure of dispersion (say the standard deviation) etc., it may not be possible to examine every element in the population. For example if we want to know about the average life time of electric bulbs produced by a particular process, if we test every element in this population, there will not be any electric bulbs left for sale. It may not be economical to conduct a survey of all citizens in India to find out the average income of the citizens. The time, money and numerous other factors involved, may compel us to adopt some other procedure in order to study the population characteristics of a given finite population. The method usually used is to take a subset (sample) from the given set (population) in such a way that the inferences or results in the subset can be generalized to the population. This is a process of induction. If our inductive inference is to be valid the subset (sample) should be a representative of the population in some sense.

If a sample is selected in such a way that every element in the population is given equal chances of being included in the sample, then such a sample is called a simple random sample from a finite population. However such a restriction on the selection is not necessary for the application of probability theory. For a detailed discussion see bibliography [5] at the end of this chapter. A simple random sample may be selected from a given population by using a table of random numbers. In such a table a set of numbers are given where the numbers from 0 to 9 occur with approximately equal proportions, the numbers from 0 to 99 occur with approximately equal proportions etc. We can number the elements in a given population of size N, from 1 to N and a simple random sample of size n can be selected by using the table of random numbers such that every element in the population is given an equal chance of being included in the sample. Such a sample may be considered to be a representative sample and the population characteristics may be estimated by some functions of the sample observations. For example the population mean may be estimated by the sample mean. The properties of such estimates will be discussed in the chapter on estimation. A random sample may be selected from a given finite population by using a random experiment. For example the elements in the population can be numbered from 1 to N. These numbers can be written on cards and cards are drawn one by one with replacement from the well shuffled deck of these N cards. Thus a sample of any required size may be obtained. A detailed discussion on the construction of random numbers see the bibliography [5], at the end of this chapter.

In some cases the given population may be divided into different strata and then a simple random sample is taken. Such sampling procedures are called stratified sampling. For example, in oder to study the average annual expenses of the families in a city the families may be divided into different income groups and simple random samples may be taken from each stratum. Sometimes sampling is done in different stages. Such sampling procedures are called multistage sampling. In order to study the attack of a particular disease in a country, a few administrative districts may be selected at random and from these districts some villages may be selected at random and survey may be conducted in these villages. Instead of taking a random sample of pre-assigned size sometimes a sequential sampling procedure is adopted. Sampling is continued or stopped based on the information obtained at every step. For a detailed discussion of these sampling procedures the reader is advised to read books on sampling.

Exercises

6.25. If an ordered set of n elements are taken at random without replacement from a set of N elements, show that the probability $f(x_1,\ldots,x_n)$ of getting n elements is,

$$f(x_1,\ldots, x_n)=1/N(N-1)\ldots(N-n+1)$$

(If all such sets of n elements have the same probability we say that $X_1,\ldots, X_n$ is a random sample from a finite population with N elements.)

6.26. Under the assumptions in problem 6·25 show that the marginal distributions, (1) $f(x_i)=1/N$ for $x_i=a_1, a_2,\ldots, a_N$ where $a_1\ldots a_N$ are the elements in the population, (2) $f(x_i, x_j)=1/N(N-1)$ where $f(x_i, x_j)$ denotes the joint probability function of X_i and X_j for some i and j,

6·27. If $\overline{X}$ is the mean of a random sample of size n selected from a finite population of size N, with mean μ and with variance σ^2, show that

$$(1)\ E(\overline{X})=\mu,\ (2)\ \text{Var}\,(\overline{X})=\frac{\sigma^2}{n}\cdot\frac{N-n}{N-1}$$

6.28. If a random sample of size n is taken with replacements from a finite population of size N with mean μ and with variance σ^2, obtain

$$(1)\ E(\overline{X}),\ (2)\ \text{Var}\,(\overline{X})$$

where $\overline{X}$ denotes the sample mean.

6.29. By using a table of random numbers select 10 samples each of size 15, 20, 30 from tbe following data and obtain the sample means. 25, 27, 26, 25, 28, 29, 26, 30, 35, 32, 36, 32, 29, 29, 31. 35, 36, 34, 32, 30, 31, 27, 28, 30, 29, 26, 30, 30, 32, 29, 32, 31, 30, 28, 29, 35, 36, 34, 32. 33, 31, 32, 30, 30, 34, 29; 31, 28, 27, 26, 29, 31, 31, 28, 34, 35, 28, 22, 23, 24, 20, 26, 28, 29, 32, 34, 36, 37, 38, 32, 20, 22, 21, 24, 23, 25, 27, 39, 40, 38, 37, 34, 30, 29, 40, 38, 22, 23, 24, 21, 20, 28, 29, 35, 36, 38, 37, 34, 36, 25, 27, 26, 23, 24, 21.

6.30. Plot the approximate distributions of the sample means of sizes $n=15$, 20, 30 in problem 6·29, by forming a frequency table of the observed sample means for the various sample sizes. Draw the three curves on the same graph.

6.31. Draw a normal curve $N(\mu, \sigma)$ with the parameter μ=the arithmetic mean of the sample means and σ^2=the variance of the sample means, of size 30 in problem 6.29. Draw this normal curve and the approximate distribution of the sample mean of size 30 obtained in problem 6.29, on the same graph.

6·32. By chosing 5 intervals of equal lengths, form a frequency table for the data in problem 6.29. Represent the frequencies by a histogram. Draw a smooth curve which is most appropriate for this histogram.

6.6. ACCEPTANCE SAMPLING

In an industrial production process where a particular article is produced in a large scale or in large number in a short interval of time, it is not practicable to examine each and every item in order to control the quality of the product. In such a situation production engineers use quality control methods and a brief account of it is given in chapter 8. In a mass production process, even if an item is produced by using the best equipments available, there will be some variations from item to item. If a particular part of a precision equipment is produced in a large scale, even the slightest departure from the specifications may make the part useless as far the consumer is concerned. In such large scale productions the items are shipped to the customer in lots of sizes, may be of thousands.

Even if a good quality control method is used, some defective items may be included in the lots. Examination of each and every item in a lot may cost more than the cost of production and if the examination can be done only by destroying the item, examination is not practicable also. If the manufacturer cannot eliminate all defectives from every lot he would like to reduce the number of defectives to a minimum and also he would like to find out the number of defectives in a lot by examining the smallest number of items possible. As a criterion he can select a random sample of size n from a lot of size N and examine the n items. If more than c of them are defective he can stop the shipment of the lot or can examine all the N items and replace the defectives by good ones. This sampling plan is based on a single sample or only one sample is taken. Such sampling plans are called single sampling plans and plans based on more than one sample are called multiple sampling plans.

6.61. The Operating Characteristic Curve. If the sampling plan is to accept a lot of size N if the number of defectives in a random sample of size n is less than or equal to c then the probability of acceptance is

$$p(\theta)=\Sigma_{x=0}^{c}\binom{N\theta}{x}\binom{N-N\theta}{n-x}\Big/\binom{N}{n}$$

where θ denotes the fraction defectives in the lot. θ can be 0 or 1/N, 2/N,...,N/N ; in other words there may be 0 or 1 or 2 or ...N defectives in the lot of size N $p(\theta)$ varies according to θ and if $p(\theta)$ is plotted against θ we get a curve called the operating characteristic curve (OC-curve) of the single sampling plan as shown in Fig. 6·1.

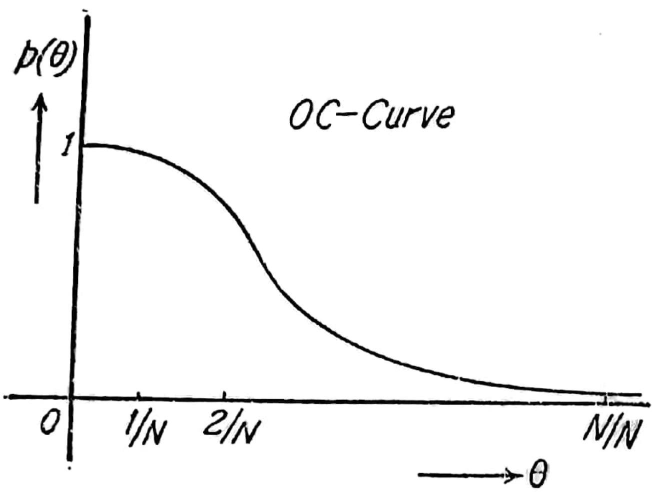


Fig. 6·1.

When N is large the hypergeometric probability $p(\theta)$ can be approximated by a Binomial probability (Ex. 4.3.1) and hence $p(\theta)$ can be approximated to.

$$p(\theta) \approx \Sigma_{x=0}^{c} \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Further when $n \to \infty$, $\theta \to 0$ but $n\theta \to \lambda$ (a constant) then $p(\theta)$ can be approximated to a Poisson probability (see section 4.24) as,

$$p(\theta) \approx \Sigma_{x=0}^{c} \frac{\lambda^x}{x!} e^{-\lambda/x\,!}$$

6.62. Producer's and Consumer's Risks. Suppose that the producer (seller) and consumer (buyer) agree to call a lot acceptable if the fraction defective $\theta \leqslant \theta_0$ and unacceptable if $\theta \geqslant \theta_1$ and also that they agree on the same sampling inspection procedure. Here θ_0 and θ_1 are known as the acceptable quality level (AQL) and lot tolerance per cent defective (LTPD) respectively. Lots with fraction defective θ is such that $\theta_0 < \theta < \theta_1$ may be called lots of indifferent quality. In this sampling inspection procedure it is possible that the producer may scrap a lot when it is really acceptable and the consumer may accept a lot when it is really unacceptable or where $\theta \geqslant \theta_1$. The probability, α, that a good lot $(\theta \leqslant \theta_0)$ is rejected is called the producer's risk and the probability, β, that a bad lot $(\theta \geqslant \theta_1)$ is accepted is called the consumer's risk. These probabilities are also known as the Type I and Type II errors respectively, in the theory of testing of hypotheses. If the seller and buyer agree on the values of α, β, θ_0 and θ_1 then at least in the large lot cases (N-large) a sampling plan can be fixed (n and c can be fixed).

Ex. 6.61.1. *In a single sampling plan, calling for a sample of size $n=20$, has the acceptance number $c=1$. Assuming that the lot size is large and a binomial approximation is appropriate, calculate the probabilities of accepting a lot when there are 10% defectives and rejecting a lot when there are only 5% defectives.*

Sol. The probability of accepting a lot when the percentage defective is 10% is,

$$=\Sigma_{x=0}^{1}\binom{20}{x}(0\cdot1)^{x}(0\cdot9)^{20-x}$$

$$=(0\cdot9)^{20}+20(0\cdot1)(0\cdot9)^{19}$$

$$=(2\cdot9)(0\cdot9)^{19}.$$

The probability of rejecting a lot when the percentage defective is 5% is,

$$=\Sigma_{x=2}^{20}\binom{20}{x}(0\cdot05)^{x}(0\cdot95)^{20-x}$$

$$=1-\Sigma_{x=0}^{1}\binom{20}{x}(0\cdot05)^{x}(0\cdot95)^{20-x}$$

$$=1-(1\cdot95)(0\cdot95)^{19}.$$

Comments. If the incoming quality of the lot is θ and if the defectives in an unacceptable lot are replaced by the non-defectives before the shipment and if $p(\theta)$ is the probability of accepting a lot then $\theta\, p(\theta)$ can be defined as the average outgoing quality (AOQ). $p(\theta)$ is the probability of accepting a lot and such lots contain proportion θ of defectives. $1-p(\theta)$ is the proportion of lots rejected in the long run.

Hence $\text{AOQ}=\theta\, p(\theta)+0[1-p(\theta)]=\theta\, p(\theta)$.

Exercises

6.33. A single sampling plan calls for $c=1$, $n=20$ when $N=40$. Draw the OC-curve. Approximate the probabilities by Binomial probabilities and compare the approximated OC-curve with the exact one.

6.34. A single sampling plan, where the lot size is large, calls for $c=2$ and $n=40$; (*a*) find AQL if the producer's risk is 0.10, (*b*) find LTPD if the consumer's risk is 0·15.

6.35. A single sampling plan calls for a sample of size 50. By using a Binomial and a Poisson approximation find, (1) the acceptance number c if AQL is 5% and the producer's risk is 0·02, (2) by using the c in (1) obtain the consumer's risk if the LTPD is 6%, (3) plot the OC curve and mark the consumer's and producer's risks.

6.36. In a single sampling plan with $n=20$ and $c=2$, plot the OC curve and AOQ curve (assume a Binomial approximation).

BIBLIOGRAPHY

1. K.A. **Brownlee,** Statistical Theory and Methodology in Science and Engineering. John Wiley, New York, 1960.

2. H.D. **Brunk,** An Introduction to Mathematical Statistics. Blaisdell Publishing Co., New York, 1965.

3. D.J. **Cowden,** Statistical Methods in Quality Control. Prentice-Hall, N.J. 1957.

4. I.G. **Guttman** & S.S. **Wilks**. Introductory Engineering Statistics. John Wiley, New York, 1965.

5. M.G. **Kendall** & A. **Stuart.** The Advanced Theory of Statistics. Vol. I. Charles Griffin, Co., London, 1958.

6. I. **Miller** & J.E. **Freund,** Probability and Statistics for Engineers. Prentice-Hall, N.J., 1965.

7. A.M. **Mood** & F.A. **Graybill,** Introduction to the Theory of Statistics. McGraw-Hill, N.Y., 1963.

8. G.H. **Weinberg** & J.A. **Schumaker,** Statistics An Intuitive Approach, Wadsworth Publishing, Co., Inc. 1965.

Tables of random numbers may be found in the following tables.

9. **D.V. Lindley and J.C.P. Miller.** Cambridge Elementary, Statistical Tables. Cambridge University Press, 1958.

10. **D.B. Owen.** Handbook of Statistical Tables. Addition Wesley Publishing Co, Inc. 1962.

CHAPTER 7

SAMPLING FROM NORMAL POPULATIONS

7.0. Introduction. In chapter 6 we have defined, a sample from a theoretical population, a statistic, sampling distributions etc. In this chapter we will study sampling from normal populations. From the central limit theorem we have seen that the normal distribution is very important in statistical analysis. In this theorem we have stated that under some general conditions such as the finiteness of the population mean and variance, the standardized sample mean [that is ; $(\bar{X}-E\bar{X})/\sqrt{\mathrm{Var}\,(\bar{X})}$] is approximately normally distributed when the sample size is large, whatever may be the population. The population from which the sample is taken need not even be continuous. There are other important results which enhance the importance of the normal distribution. Some such results are mentioned in problem 7.30. Further in many practical problems, where some general conditions are satisfied it can be shown that a normal distribution is a good fit to the data under consideration In some non-normal cases appropriate transformations exist by which we can make the transformed variable a normal variable. So samples and sampling distributions, from normal populations, play a vital role in statistical theory, especially in testing statistical hypotheses.

7.1. A SAMPLE FROM A NORMAL POPULATION

If the stochastic variables $X_1, \ldots, X_n$ are independently and identically distributed as a $N(\mu, \sigma)$ then we say that $X_1, \ldots, X_n$ is a simple random sample from a $N(\mu, \sigma)$ or $x_1. \ldots, x_n$ is called an observed random sample from a $N(\mu, \sigma)$. (A numerical sample is taken as an observed sample).

7.2. THE DISTRIBUTION OF THE SAMPLE MEAN

If $X_1, \ldots, X_n$ is a simple random sample from a $N(\mu, \sigma)$ then $\bar{X}=(X_1+\ldots+X_n)/n$ has a normal distribution with the mean μ and with variance σ^2/n. This can be seen from Ex. 6.1.1. Hence the standardized variable $Y=(\bar{X}-\mu)/(\sigma/\sqrt{n})$ has a standard normal distribution. The density function for $\bar{X}$ may be given as,

$$f(\bar{x}) = (n/2\pi\sigma^2)^{1/2} e^{-n(\bar{x}-\mu)^2/2\sigma^2} \quad -\infty<\bar{x}<\infty,$$
$$-\infty<\mu<\infty,\ \sigma>0. \tag{7.1}$$

Evidently $E(\bar{X})=\mu$ and $\mathrm{Var}\,(\bar{X}) = \sigma^2/n$. Therefore probabilities like $P\{\bar{x}>c\}$, $P\{c<\bar{x}<d\}$ etc., may be evaluated by using a normal probability table. We have seen that, if the parent population is normal with the parameters μ and σ, then the statistic, sample mean, has a normal distribution with the parameters μ and $\sigma/\sqrt{n}$.

Ex. 7.2.1. *A dressmaker made the following observations. The waist measurements of 9 girls of a particular age group give an average of 20″. If he has enough evidence to justify his assumption that the waist measurements in that age group are distributed as a $N(\mu=25, \sigma=2)$, what is the probability that, in the long-run, he gets an average greater than 26″. If he is taking measurements for batches of 9 girls ?*

Sol. According to our notation $\mu=25$, $\sigma=2$, $n=9$. The sample mean has a normal distribution with the parameters $\mu=25$ and $\sigma/\sqrt{n}=2/3$. That is,

$$f(\bar{x})=(3/2\sqrt{2\pi})\ e^{-9(x-25)^2/4}$$

The required probability=

$$P\{\bar{x}\geqslant 26\}=\int_{26}^{\infty} f(\bar{x})\ d\bar{x}.$$

Let $\qquad t=(\bar{x}-25)/(2/3)$

then $\qquad 2\ dt/3=d\bar{x}$ and when $\bar{x}=26$, $t=3/2$.

$$\text{Therefore,} \int_{26}^{\infty} f(\bar{x})\ d\bar{x}=\int_{3/2}^{\infty} (2\pi)^{-1/2}\ e^{-t^2/2} dt=0{\cdot}0668.$$

(from normal tables)

Ex. 7.2.2. *A man fishing at a particular spot in the Vembanad lake in Kerala caught 16 fish and the average length was 12″. Assuming that these length measurements are a random sample from a $N(\mu, \sigma=2)$ find two quantities t_0 and t_1 such that we can say with a probability of 0·95 that the expected length μ lies between t_0 and t_1 or*

$$P\{t_0<\mu<t_1\}=0{\cdot}95.$$

Sol. We can find out t_0 and t_1 by using the property that,

$$Y=(\bar{X}-\mu)/(\sigma/\sqrt{n}) : N(0, 1).$$

In other words, Y has a density function,

$$f(y)=\frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}, -\infty<y<\infty$$

and the distribution is given in Fig. 7.1.

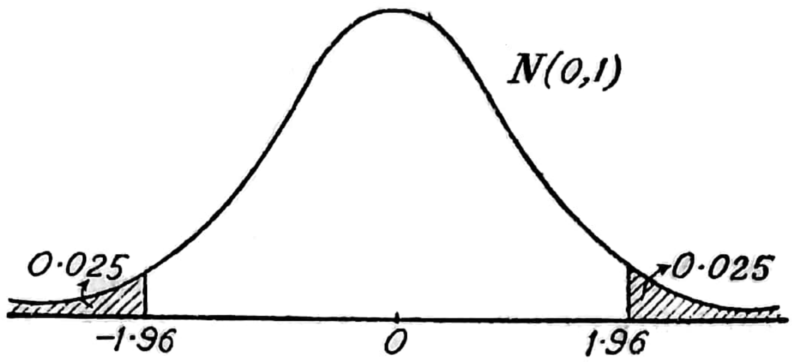


Fig. 7.1.

From normal probability tables it is seen that

$$\int_{1.96}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}dy = 0{\cdot}025 \tag{7.2}$$

Therefore if we take $y_0=-1.96$ and $y_1=1.96$, we have

$$P\{y_0 \leqslant y \leqslant y_1\}=0.95$$

$$P\left\{-1.96 \leqslant \frac{(\bar{x}-\mu)}{\sigma/\sqrt{n}} \leqslant 1.96\right\}=0.95$$

$$P\left\{-1.96\frac{\sigma}{\sqrt{n}} \leqslant \bar{x}-\mu \leqslant 1.96\frac{\sigma}{\sqrt{n}}\right\}=0.95$$

or $$P\left\{\bar{x}-1.96\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{x}+1.96\frac{\sigma}{\sqrt{n}}\right\}=0.95$$

Therefore, $P\{12-1.96)(2/4) \leqslant \mu \leqslant 12+1.96(2/4)\}=0.95$.

Hence, $t_0=12-1.96(2/4)=11.02$ and $t_1=12+0.98=12.98$.

Ex. 7.2.3. *A random sample of 25 pepper plants from a pepper plantation yield an average of 25 lbs. of pepper in a particular year Assuming that the distribution of yields is a $N(\mu, \sigma=4)$ will you accept the hypothesis that the expected yield per plant, that is, $\mu=30$ lbs. Suppose that we are ready to accept the hypothesis if the probability of getting a sample mean as small as the observed one is at least 0.05.*

Sol. $$t=(\bar{X}-\mu)/(\sigma/\sqrt{n})\ ;\ N(0, 1).$$

If $\mu=30$ then $(\bar{x}-\mu)/(\sigma/\sqrt{n})=25(25-30)/4=-6{\cdot}25$.

The probability of getting a t as small as -6.25 is,

$$= \int_{-\infty}^{-6.25} (1/2\pi)^{1/2} \; e^{-t^2/2}\, dt < 0.03$$

(from normal tables.)

Hence we reject the hypothesis.

Exercises

7.1. A random sample of size 20 is taken from a $N(\mu, \sigma)$ where $\mu=30$ and $\sigma=4$. What is the probability that the sample mean will fall between 25 and 35 ?

7.2. A random sample of size, 15 is taken from a $N(\mu, \sigma)$ where $\sigma=2$. What is the probability that $3(\bar{x}-\mu) \geqslant 4$ where $\bar{x}$ denotes the sample mean ?

7.3. A random sample of size 20 is taken from a $N(\mu, \sigma)$ where $\sigma=5$, What is the probability that the sample mean will not differ from the population mean by more than 2 in absolute value ?

7.4. A random sample of size 40 from a $N(\mu, \sigma)$ where $\sigma=4$, has a mean 35. Find two quantities t_0 and t_1 such that $P\{t_0 \leqslant \mu \leqslant t_1\}=0{\cdot}99$. Are t_0 and t_1 unique ?

7.5. A random sample of size 30 from a $N(\mu, \sigma)$ where $\sigma=25$, has a mean 42.5, Find four quantities t_0, t_1, t'_0, t'_1 such that $P\{t_0 \leqslant \mu \leqslant t_1\}=0.95$ $=P\{t'_0 \leqslant \mu \leqslant t'_1\}$.

7.6. Two independent random samples of sizes 20 and 25 are taken from two normal populations $N(\mu, \sigma_1=2)$ and $N(\mu, \sigma_2=5)$ respectively. What is the probability that the sample means will not differ by more than 3 in absolute value ?

7.7. Two independent random samples of sizes 15 and 20 from $N(\mu_1, \sigma_1=2)$ and $N(\mu_2, \sigma_2=3)$ have means 30 and 32, respectively. Find t_0 and t_1 such that $P\{t_0 \leqslant \mu_1-\mu_2 \leqslant t_1\}=0{\cdot}95$. Are t_0 and t_1 unique ?

7.8. The average yield of corn in 10 experimental plots is 20 bushels. If the distribution of yield is $N(\mu, \sigma)$ with the true average yield μ and with a standard deviation $\sigma=4$, will you accept the hypothesis that the true average yield is 18, assuming that we are ready to accept the hypothesis if the probability of getting the sample mean as large as the observed mean, is at least 0.05.

7.9. Two independent random samples of 10 boys and 15 girls have average I.Q's 104 and 103 respectively. If the I.Q's are distributed as $N(\mu_1, \sigma_1=4)$ and $N(\mu_2, \sigma_2=4)$, will you accept the hypothesis that boys and girls are equally intelligent, taking the same acceptance level as in problem 7.8.

7.3. THE CHI-SQUARE (χ^2) DISTRIBUTION

Another important sampling distribution is the chi-square distribution. If X: $N(0, 1)$ or if X has a standard normal distribution X^2 has a gamma distribution with the parameters $\alpha=1/2$ and $\beta=2$. This was seen in problem 4.46 of chapter 4. If we consider the independent *s.v*'s, $X_1, X_2, \ldots X_k$ where $X_i: N(0, 1)$ for $i=1, 2, \ldots k$, then the statistic $X_1^2+X_2^2+\cdots+X_k^2$ is called a χ^2 statistic with k degrees of freedom. In other words a χ^2 statistic with k degrees

of freedom is the sum of squares of k independent standard normal variates. This χ^2 statistic has a χ^2 distribution and the density function is given by

$$f(x)=\frac{1}{2^{k/2}\Gamma(k/2)}\,x^{\frac{k}{2}-1}\,e^{-x/2},\ 0<x<\infty \tag{7.3}$$

(k—a positive integer)

$=0$ elsewhere,

where k, the degrees of freedom, is the only parameter. This distribution may be derived as follows.

Let $$\chi^2=Y_1+Y_2+\ldots+Y_k \tag{7.4}$$

where $Y_i=X_i^2$ and X_i : N(0, 1) for $i=1, 2, \ldots k$ and all the X's are independent.

Y_i has a Gamma distribution with the parameters $\alpha=1/2$ and $\beta=2$ for all $i=1, 2,\ldots k$. The M.G.F. for Y_i is

$$M_{Y_i}(t)=(1-2t)^{-1/2}$$

$$\therefore \quad M\chi^2(t) = \prod_{i=1}^{k}(1-2t)^{-1/2}=(1-2t)^{-k/2}. \tag{7.5}$$

By the uniqueness property of the moment generating functions, χ^2 has a Gamma distribution with the parameters $\alpha=k/2$ and $\beta=2$.

Hence the density function is as given in (7.3).

The χ^2 variate with k degrees of freedom has a moment generating function

$$M^2\ (t)=(1-2t)^{-k/2} \tag{7.6}$$

The distribution is given in Fig. 7.2. The shape of the curve depends upon the degrees of freedom k. For $k=2$ it is the exponential distribution.

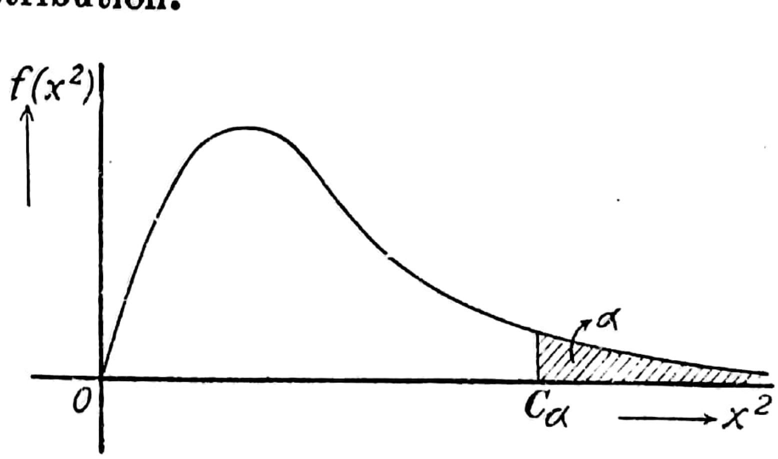


Fig. 7.2.

Let $\quad P\{\chi^2 \geqslant c_\alpha\} = \alpha = \int_{c_\alpha}^{\infty} f(\chi^2)\, d\chi^2$

where C_α is a point such that, the tail area from C_α to ∞ or the area under the curve between the ordinates at $\chi^2 = C_\alpha$ and $\chi^2 = \infty$, is α. For any given α and the degrees of freedom k we can find out the C_α. This C_α is tabulated for various values of α and the degrees of freedom k. Such a table is called a χ^2 table. An extract is given at the end of this book. A χ^2 distribution has many practical applications which we will see from the following results.

Theorem 7.1. If X_1 and X_2 are independently distributed as χ^2 variates with k_1 and k_2 degrees of freedom respectively then $Y = X_1 + X_2$ has a χ^2 distribution with $k_1 + k_2$ degrees of freedom.

Proof. The M.G.F. for X_1 is

$$M_{X_1}(t) = (1-2t)^{-k_1/2}$$

and

$$M_{X_2}(t) = (1-2t)^{-k_2/2}$$

$$M_Y(t) = M_{X_1}(t)\, M_{X_2}(t)$$

(since X_1 and X_2 are independent)

$$= (1-2t)^{-(k_1+k_2)/2} \tag{7.7}$$

But $(1-2t)^{-(k_1+k_2)/2}$ is the M.G.F. of a χ^2 variable with $k_1 + k_2$ degrees of freedom. Therefore from the uniqueness property of M.G.F's, Y has a χ^2 distribution with the parameter $(k_1 + k_2)$.

Corollary 1. If $X_1, X_2, \ldots X_n$ are n independent χ^2 variables with $k_1, k_2, \ldots, k_n$ degrees of freedom respectively, then $Y = X_1 + X_2 + \ldots + X_n$ is a χ^2 variable with $k_1 + k_2 + \ldots + k_n$ degrees of freedom.

Corollary 2. If X_1 and X_2 are two independent stochastic variables where X_1 has a χ^2 distribution with k_1 degrees of freedom and X_1, X_2 has a χ^2 distribution with k_2 degrees of freedom $(k_2 > k_1)$, then X_2 has a χ^2 distribution with $k_2 -_1 k$ degrees of freedom.

A χ^2 variable with k degrees of freedom is often denoted by χ^2_k where the suffix k denotes the degrees of freedom. From the definition of the χ^2 variate it may be noted that k is a positive integer. We can find out a number of statistics which have a χ^2 distributions. For example consider a random sample $X_1, X_2, \ldots X_n$ from a normal population $N(\mu, \sigma)$ then the statistic

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \frac{(X_1-\mu)^2}{\sigma^2} + \frac{(X_2-\mu)^2}{\sigma^2} + \ldots + \frac{(X_n-\mu)^2}{\sigma^2} \tag{7.8}$$

is evidently the sum of squares of n independent standarized normal variables and hence it is a χ^2 with n degrees of freedom. Further,

$$\sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma^2} = \frac{1}{\sigma^2}\left\{\sum_{i=1}^{n} (X_i-\overline{X}+\overline{X}-\mu)^2\right\}$$

$$=\frac{1}{\sigma^2}\left\{\sum_{i=1}^{n} (X_i-\overline{X})^2+n(\overline{X}-\mu)^2\right\}$$

$$=\sum_{i=1}^{n} \frac{(X_i-\overline{X})^2}{\sigma^2}+n\frac{(\overline{X}-\mu)^2}{\sigma^2}$$

$$=n\, S^2/\sigma^2+(\overline{X}-\mu)^2/(\sigma/\sqrt{n})^2 \tag{7.9}$$

where $S^2=\sum_{i=1}^{n} (X_i-\overline{X})^2/n=$ the sample variance.

But $E(\overline{X})=\mu$ and Var $(\overline{X})=\sigma^2/n$. Hence $(\overline{X}-\mu)^2/(\sigma^2/n)$ is a χ^2 with one degree of freedom. By corollary 2 of theorem 7.1, nS^2/σ^2 is a χ^2 variable with $n-1$ degrees of freedom, if we can show that nS^2/σ^2 and $(\overline{X}-\mu)^2/(\sigma^2/n)$ are independent. This will not be proved here (see reference 3 and problem 7.32). We will state the following theorem without proof.

Theorem 7.2. If $X_1, X_2, \ldots, X_n$ is a random sample from a normal population $N(\mu, \sigma)$ and if

$$S^2=\Sigma_{i=1}^{n} (X_i-\overline{X})^2/n$$

is the sample variance, then

$$nS^2/\sigma^2=\Sigma_{i=1}^{n} (X_i-\overline{X})^2/\sigma^2$$

has a χ^2 distribution with $n-1$ degrees of freedom (d. f.) (7.10)

When the degrees of freedom of a χ^2 variate is sufficiently large the χ^2 distribution approximates to a normal distribution. When the degrees of freedom $k>30$ a good normal approximation may be obtained. If $X_1, \ldots X_k$ are k independent normal variables with $E(X_i)=\mu$ for all i and Var $(X_i)=1$ for all i then the distribution of $Y=X_1^2+\ldots+X_k^2$ is called a non-central χ^2 distribution.

Ex. 7.3.1. *An endurance test is conducted on a random sample of 10 persons from a city, to study the ability to stand pain. The sample variance is found to be 16 units. If such batches of 10 are tested what is the probability that the experimenter gets a sample variance as large as 18, assuming that the endurance measurements are distributed as a $N(\mu, \sigma=3)$?*

Sol. According to our notation $n=10$, $\sigma=3$. But nS^2/σ^2 is with $n-1=9$ d.f. $nS^2/\sigma^2=10\times18/9=20$ when $s^2=18$.

Therefore the required probability $=P\{\chi_9^2 \geqslant 20\}$. From the χ^2 tables,

$$\int_{20}^{\infty} f\left(\chi_9^2\right)d\chi_9^2=0.02 \text{ approximately.}$$

Hence the required probability $=0{\cdot}02$ approximately.

Ex. 7.3.2. *A random sample of 11 metal rods produced by a particular process are tested for breaking strength and it is found that the sample variance is 16 units. Find out two quantities t_0 and t_1 such that we can say with a confidence of 95% that the unknown σ^2 will lie between t_0 and t_1 or $P\{t_0 \leqslant \sigma^2 \leqslant t_1\}=0.95$.*

Sol. nS^2/σ^2 is a χ^2 with $n-1$ d.f.

Let y_0 and y_1 be such that

$$\int_0^{y_0} f\left(\chi_{10}^2\right)dx_{10}^2=0{\cdot}025=\int_{y_1}^{\infty} f\left(\chi_{10}^2\right)d\chi_{10}^2$$

This is shown in Fig. 7.3.

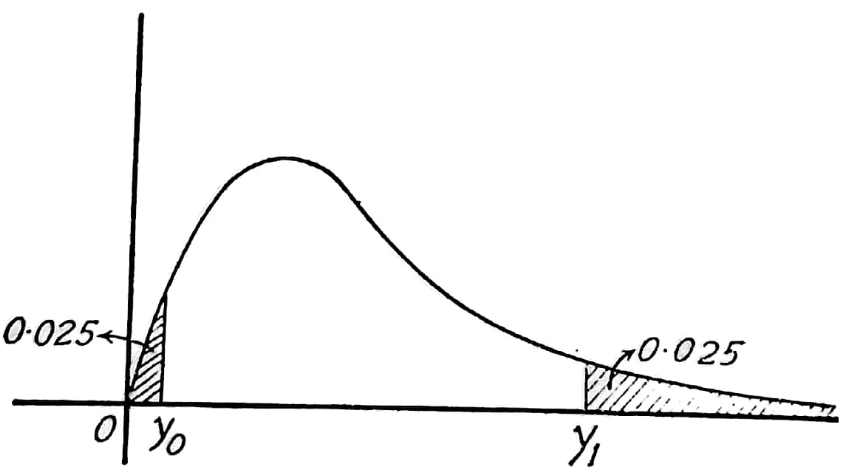


Fig. 7.3.

where $f\left(\chi_{10}^2\right)$ is the density function of a χ^2 with 10 d.f.

$$\therefore \qquad P\{y_0 \leqslant 11\, s^2/\sigma^2 \leqslant y_1\}=0.95$$

$$P\left\{\frac{y_0}{11.s^2} \leqslant \frac{1}{\sigma^2} \leqslant \frac{y_1}{11.s^2}\right\}=0.95$$

$$P\left\{\frac{11.s^2}{y_1} \leqslant \sigma^2 \leqslant \frac{11.s^2}{y_0}\right\}=0.95$$

$$\left(\text{If } a<b \text{ then } \frac{1}{a}>\frac{1}{b}\right)$$

Here $s^2=16$ and from chi-square tables $y_0=3.247$

and $y_1=20.483$,

Therefore, $t_0=11\ s^2/y_1=(11)(16)/20.483=8.6$

and $t_1=11\ s^2/y_0=(11)(16)/3.247=54.2.$

Ex. 7.3.3. *A random sample of 15 fish who could jump over a dam in a river were caught and weighed. A sample variance of 10 units is observed. Assuming that the weights of those fish who could jump over the dam has a $N(\mu, \sigma)$, will you accept the hypothesis that $\sigma=6$? Suppose that we will accept the hypothesis if the probability of getting a chi-square as large as an observed one is at least 0.10.*

Sol. nS^2/σ^2 is a chi-square with $n-1$ degrees of freedom. If $\sigma^2=6$ then $ns^2/\sigma^2=15(10)/6=25$ is an observed value of a χ^2 with 14 degrees of freedom. The probability of getting an observed χ^2 as large as 25 is,

$$\int_{25}^{\infty} f\left(\chi^2_{14}\right) d\chi^2_{14} < 0.10 \text{ (from chi-square tables).}$$

Hence the hypothesis is rejected.

Exercises

7.10. If X_1, X_2 is a random sample of size 2 from a N(0, 1), show that the sample mean and the sample variance are independently distributed.

7.11. If X is a χ^2 with n degrees of freedom, show that $\sqrt{2X}-\sqrt{2n}$ is approximately normally distributed when n is sufficiently large.

7.12. A random sample of size 10 is taken from a $N(\mu, \sigma)$ where $\sigma=5$. What is the probability that the sample variance is as large as 36 or as small as 9 ?

7.13. A random sample of size 20 is taken from a $N(\mu, \sigma)$ where $\sigma=10$. What is the probability that the ratio of the sample variance to the population variance will not exceed unity ?

7.14. A random sample of size 15 from a $N(\mu, \sigma)$ has a variance 16. Find out two quantities t_0 and t_1 such that $P\{t_0 \leqslant \sigma^2 \leqslant t_1\}=0 \cdot 95$. Are t_0 and t_1 unique ?

7.15. Paper bags are filled up with peanuts by a machine. A random sample of 4 such bags yield the following data. $\Sigma x_i=20$, $\Sigma x_i^2=120$, where x_i is the weight of the i^{th} packet. Assuming that the distribution of the weights is normal what is the probability that the sample came from a population with $\sigma^2=0.1$?

7.4. THE STUDENT DISTRIBUTION

In section 7.2 it is seen that the statistic $Y=(\overline{X}-\mu)/(\sigma/\sqrt{n})$ is a standard normal variable if $X_1, \ldots, X_n$ is a random sample from a normal population $N(\mu, \sigma)$ and if $\overline{X}$ is the sample mean. But

often the population standard deviation σ is unknown. In that case we would like to consider some other statistic which does not contain σ. The distribution of the new statistic may be of some use to us in doing some problems or testing some hypotheses etc. we know that

$$E\frac{\Sigma(X_i-\bar{X})^2}{n-1}=\sigma^2 \tag{7·11}$$

If we replace σ in Y by the square root of this unbiased estimator for σ^2 then we get a new statistic called the student t statistic. That is the student t statistic is

$$t=(\bar{X}-\mu)/(S'/\sqrt{n}) \tag{7.12}$$

where $$S'^2=\sum_{i=1}^{n}(X_i-\bar{X})^2/(n-1)$$

The distribution of this statistic t is called the student t distribution. This distribution was first given by W.S. Gosset who used the pen name student and hence the distribution is called a student t distribution. We know that

$$\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{\sigma^2}=\frac{(n-1)}{\sigma^2}S^2$$

has a χ^2 distribution with $n-1$ degrees of freedom. Hence

$$t=(\bar{X}-\mu)/(S'/\sqrt{n})$$

is called a student t with $n-1$ degrees of freedom, and is usually denoted by t_{n-1}. The density function of a student t with v degrees of freedom is.

$$f(t)=\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v}\sqrt{\pi}\Gamma(v/2)}(1+t^2/v)^{-(v+1)/2}$$

$$\text{for } -\infty<t<\infty,\ v>0 \tag{7.13}$$

It is easy to derive this distribution since,

$$t_{n-1}=(\bar{X}-\mu)/(S'/\sqrt{n})=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\bigg/\frac{S'}{\sigma} \tag{7.14}$$

The numerator of t_{n-1} is a standard normal variate and the denominater is a $\left(\chi^2_{n-1}/(n-1)\right)^{1/2}$, since $(n-1)S'^2/\sigma^2$ is a χ^2_{n-1}. Further the numerator and the denominator are independent. This is mentioned in section 7.3 and also see problem 7.32. This independence of the sample mean and the sample variance is a characteristic property of the normal distribution. That is, the

sample mean and the sample variance are independently distributed if and only if the population is normal. For a proof of this result and for other characteristic properties of the normal distribution see reference 3 at the end of this chapter. In order to obtain the distribution of a student t with v degrees of freedom we will consider the distribution of $t=X/\sqrt{Y/v}$ where X and Y are independent, X is a standard normal variate and Y is a χ^2, variate with v degrees of freedom. The joint density function of X and Y is,

$$f(x, y)=(2\pi)^{-1/2}\;{}^{-x^2/2}\frac{1}{2^{v/2}\Gamma(v/2)}\,y^{v/2-1}\,e^{-y/2} \qquad (7.15)$$

$$\text{for } y>0,\ -\infty<x<\infty$$

$$=0 \text{ elsewhere.}$$

Let us consider the transformation,

$$t=x/\sqrt{y/v}$$

$$u=y \qquad (7.16)$$

The Jacobian of the transformation

$$J=\begin{vmatrix} \dfrac{\partial t}{\partial x} & \dfrac{\partial u}{\partial x} \\[2ex] \dfrac{\partial t}{\partial y} & \dfrac{\partial u}{\partial y} \end{vmatrix}$$

$$=v^{\frac{1}{2}}\;y^{-\frac{1}{2}}$$

$$f(t, u)=f(x, y)\;v^{-\frac{1}{2}}\;y^{\frac{1}{2}}$$

$$=\frac{v^{-\frac{1}{2}}}{\sqrt{2\pi}\;2^{v/2}\Gamma(v/2)}\,u^{(v-1)/2}\,e^{-\frac{u}{2}\left(1+\frac{t^2}{v}\right)}$$

$$\text{for } -\infty<t<\infty,\ u>0$$

$$=0 \text{ elsewhere.} \qquad (7{\cdot}17)$$

Now integrating out u we get

$$f(t)=\frac{\Gamma\left(\dfrac{v+1}{2}\right)}{\sqrt{v}\,\sqrt{\pi}\,\Gamma(v/2)}\left(1+\frac{t^2}{v}\right)^{-(v+1)/2} \qquad (7.18)$$

$$\text{for } -\infty<t<\infty$$

Evidently $f(t)$ is symmetric about the $f(t)$-axis and t_α for which

$$\int_{t_\alpha}^{\infty} f(t)\,dt=\alpha$$

is tabulated for various values of α and for various values of the degrees of freedom v. Such a table is called a student t table. An extract is given at the end of this book. The student t distribution is given in Fig. 7.4.

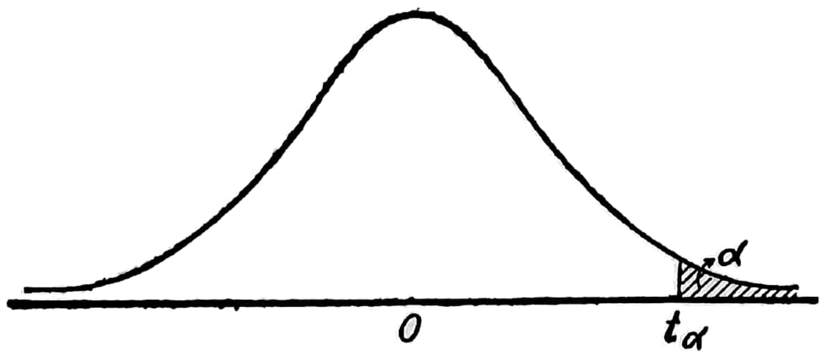


Fig. 7.4.

When the degrees of freedom v is sufficiently large, the student t distribution approximates to a normal distribution. A good approximation is obtained when $v \geqslant 30$. If instead of $t=(\bar{X}-\mu)/(S'/\sqrt{n})$ we take $Y=\bar{X}/(S'/\sqrt{n})$ then Y is called a non-central t and its distribution is known as a non-central t distribution.

Ex. 7.4.1. *Bags are filled with an expected weight of 20 lbs of potatoes by an automatic device which can only count the potatoes. A random sample of 9 bags shows an average weight of 18 lbs and $s'^2=\Sigma(x_i-\bar{x})^2/(n-1)=16$ lbs. If samples of 9 bags are taken, what is the probability that one gets a sample with an average weight exceeding 22 lbs., assuming that the weight distribution is a $N(\mu=20, \sigma)$?*

Sol. Here the population standard deviation σ is unknown ; $t=(\bar{X}-\mu)/(S'/\sqrt{n})$ has a student t distribution with $n-1$ d.f. According to our notation $n=9$, $\mu=20$, $s'=4$. When $\bar{x}=22$, $t=(22-20)/(4/\sqrt{9})=3/2$. Hence the required probability is,

$$P\{t \geqslant 3/2\}=\int_{3/2}^{\infty} f(t)\, dt \text{ and the d.f.}=n-1=8.$$

From t-tables $P\{t \geqslant 3/2\}=0{\cdot}085$ approximately.

Ex. 7.4 2. *A random sample of size 20 from a $N(\mu, \sigma)$ has a sample mean 25 and a sample variance 16. Will you accept the hypothesis that $\mu=24$. (Suppose that we are ready to accept the hypothesis if the probability of getting a student t as large as the observed one, is at least $0{\cdot}05$).*

Sol. We know that $t=(\bar{X}-\mu)/(S'/\sqrt{n})$

where $$S'^2=nS^2/(n-1),$$

is a student t with $n-1$ degrees of freedom.

Here $$n=20,\ s^2=16.$$

$\therefore \qquad ns^2/(n-1)=20(16)/19=17$ approximately.

If $\qquad \mu=24$, then

$$(\bar{x}-\mu)/(s'/\sqrt{n})=(25-24)\sqrt{20}/\sqrt{17}$$

$$=1{\cdot}09 \text{ approximately.}$$

If our hypothesis $\mu=24$, is correct $1{\cdot}09$ is an observed value of a student t with 19 degrees of freedom.

The probability of getting a t as large as 1.09

$$\int_{1.09}^{\infty} f(x)\, dx.$$

where $f(x)$ is the density function of a student t with 19 degrees of freedom.

From the student t tables

$$\int_{1.09}^{\infty} f(x)\, dx > 0{\cdot}05$$

Here we will accept the hypothesis.

Exercises

7.16. A random sample of size 20 from a $N(\mu, \sigma)$ has a variance 18. If μ is estimated by the sample mean $\bar{X}$ then $|\bar{X}-\mu|$ may be called the error in the estimation. Find the probability that this error will not exceed 2.

7.17. A random sample of 10 citizens in a big city have an average income of $10,000 with a standard deviation of $200. If the income distribution of the citizens in the city is approximately $N(\mu, \sigma)$, obtain a 95% interval estimate for μ or in other words obtain t_0 and t_1 such that

$$P\{t_0 \leqslant \mu \leqslant t_1\}=0.95.$$

7.18. A random sample of 20 university students have an average height of 65″ with a standard deviation of 2″. If the height measurements of the university students under consideration is assumed to be approximately $N(\mu, \sigma)$, is it reasonable to take a decision that $\mu=64''$?

7.19. Obtain $E(t)$ and Var (t) where t is a student t with n degrees of freedom.

7.20. If two independent random samples of sizes 20 and 25 from $N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$ have variances 16 and 18 respectively, is it reasonable to take a decision that $\mu_1=\mu_2$, based on the above observations ?

7.21. Two independent random samples of sizes 10 and 12 are taken from a $N(\mu, \sigma_1=2)$ and a $N(\mu, \sigma_2=5)$ respectively. By using Chebyshev's inequality or otherwise obtain a probability limit that $\bar{X}_1 \geqslant \bar{X}_2+5$ where $\bar{X}_1$ and $\bar{X}_2$ are the sample means of the two samples.

7.22. Two independent random samples of sizes 20 and 30 are taken from a $N(\mu=50, \sigma=4)$ and a $N(\mu=45, \sigma=3)$ respectively. Obtain a probability bound that $\bar{X}_1$ will not differ from $\bar{X}_2$ by more than 3 units, where $\bar{X}_1$ and $\bar{X}_2$ denote the sample means.

7.23. A random sample of size 15 is taken from a population with mean μ and standard deviation $\sigma=6$. Find out k such that the probability that $2\,|\bar{X}-\mu| \leqslant k$ is at least 0·99, where $\bar{X}$ is the sample mean.

7.24. A machine part with a specified diameter of 10 units is produced by a production process. The diameters of a random sample of 5 are 10.001, 10.002, 9.99, 9.98, 10.01. If such random samples of size 5 are taken what is the probability that 2 out of 3 random samples of size 5 will have the average diameter between 9.99 and 10.01 ?

[**Hint :** Obtain $P\{9.99 \leqslant x \leqslant 10.01\}$; then apply the Binomial probability law.]

7.5. THE F-DISTRIBUTION

We have seen several important statistics like the χ^2 statistic, the student t statistic etc. Now we will define an F statistic. If we have two independent χ^2 statistics with m and n degrees of freedom respectively, then the ratio

$$F_{m,\,n}=\frac{\chi^2_m/m}{\chi^2_n/n} \tag{7.19}$$

is called an F statistic and its distribution is called an F-distribution. This distribution is obtained in Ex. 5.5.3. It is also obtained as the distribution of a transformed beta variable in problem 5.31 of Chapter 5. If we consider a student t statistic with v degrees of freedom that is,

$$t_v=X/\sqrt{\chi^2_v/v}$$

where $X : N(0, 1)$ and χ^2_v is a χ^2 with v degrees of freedom, t^2 is evidently an F-statistic. The numerator of t^2_v is a $\chi^2_1/1$ and the denominator is a χ^2_v/v and further these two χ^2 's are independent by the assumptions in t_v. Since there are two degrees of freedom attached to an F-statistic (that is, the degrees of freedom for the numerator χ^2 and the degrees of freedom for the denominator χ^2) we always say an F with m and n degrees of freedom, where m is the degrees of freedom for the numerator χ^2 and n is the degrees of freedom for the denominator χ^2.

For example t^2_v which is mentioned above is an F with 1 and v degrees of freedom. We will use the notation $F_{m,\,n}$ (F with

m and n degrees of freedom). In section 7.3 we have seen that if a random sample of size n from a $N(\mu, \sigma)$ has a mean $\overline{X}$ and a sample variance S^2 then nS^2/σ^2 is a χ^2 with $n-1$ degrees of freedom. Now we can state the following theorem.

Theorem 7.3. If two independent random samples of sizes n_1 and n_2 from two normal populations $N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$, have the sample variances S_1^2 and S_2^2 respectively, then

$$\frac{n_1 S_1^2/(n_1-1)}{n_2 S_2^2/(n_2-1)} = F_{n_1-1,\, n_2-1} \tag{7.20}$$

or in other words, $\left[n_1 S_1^2/(n_1-1) \right] \Big/ \left[(n_2 S_2^2/(n_2-1) \right]$ has an F-distribution with n_1-1 and n_2-1 degrees of freedom.

Proof. If $X_1, X_2, \ldots X_{n_1}$ and $Y_1, Y_2, \ldots Y_{n_2}$ denote the two samples

$$\frac{n_1 S_1^2}{\sigma^2} = \Sigma \frac{(X_i - \overline{X})^2}{\sigma^2} = \chi^2_{n_1-1}$$

and $$n_2 S_2^2 /\sigma^2 = \sum_{i=1}^{n} (Y_i - \overline{Y})^2/\sigma^2 = \chi^2_{n_2-1}$$

Hence $$\frac{\left[n_1 S_1^2 / \sigma^2 \right] \Big/ \left(n_1-1\right)}{\left[n_2 S_2^2 / \sigma^2 \right] \Big/ \left(n_2-1\right)} = \frac{n_1 S_1^2 \Big/ \left(n_1-1\right)}{n_2 S_2^2 \Big/ \left(n_2-1\right)}$$

$$= F_{n_1-1,\, n_2-1}$$

(Since the two χ^2 's are independent by the assumptions in the theorem).

This interesting result makes the F-distribution an important sampling distribution. This distribution is sometimes called the variance-ratio distribution. The density function of an $F_{m,\, n}$ is given by

$$f(x, \theta) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{n}\right)} \left(\frac{m}{n}\right)^{m/2} \frac{x^{\frac{m}{2}-1}}{\left(1+\frac{m}{n}x\right)^{(m+n)}}$$

for $0 < x < \infty$

$\theta = (m, n)$-positive integers. (7.21)

$= 0$ elsewhere.

A graphical representation is given in Fig. 7.5. The shape of the curve varies with the degrees of freedoms m and n.

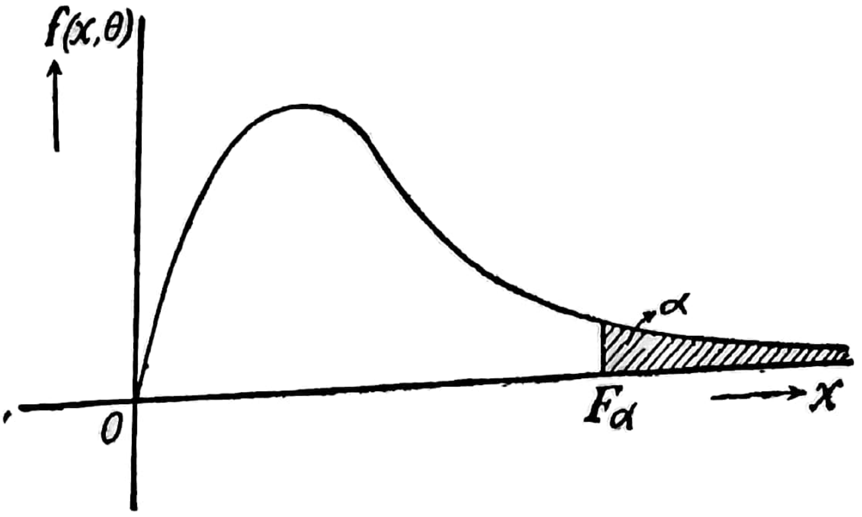


Fig. 7.5.

Here m and n are the parameters. The tail area, as shown in the figure is,

$$\alpha = \int_{F_\alpha}^{\infty} f(x, \theta)\, dx$$

F_α for various values of the degrees of freedoms m and n and for various values of α, is tabulated. Such tables are called F-tables. An extract is given at the end of this book. If the numerator χ^2 in an F is a non-central χ^2 then such an F is called a non-central F and its distribution is called a non-central F distribution.

Ex. 7.5.1. *Two random samples of 25 and 26 students are taken from students who are interested in higher altitude flying, and their heights are measured. What is the probability that the ratio of the sample variances (first to the second) is at least 3, assuming that the height measurements of such students has an approximate normal distribution ?*

Sol. Here the samples are from the same Normal population. Hence,

$$[n_1 S_1^2/(n_1-1)]/[n_2 S_2^2/(n_2-1)] = S_1'^2 / S_2'^2 = F_{n_1-1,\, n_2-1}$$

or
$$\frac{n_1}{n_2}\,\frac{n_{2}-1}{n_1-1}\,\frac{S_1^2}{S_2^2} \text{ is } F_{n_1-1,\, n_2-1}$$

Here $n_1=25$ and $n_2=20 \Rightarrow \dfrac{n_1}{n_2}\,\dfrac{n_2-1}{n_1-1}$

$=0.99$ approximately

$$\therefore \; P\left\{\frac{S_1^2}{S_2^2} \geqslant 3\right\} = P\left\{\frac{n_1}{n_2}\,\frac{(n_2-1)}{(n_1-1)}\,\frac{S_1^2}{S_2^2} \geqslant (0.99)(3)\right\} = 2.97$$

$$= P\left\{F_{24,\,19} \geqslant 2.97\right\}$$

$$= \int_{2.97}^{\infty} f(x)dx = 0.01 \text{ (approximately)}$$

where $f(x)$ is the density function for an F with 24 and 19 degrees of freedom, and the probability 0.01 is obtained from an F-table.

Ex. 7.5.2. *Two random samples of sizes 15 and 10 from two normal populations $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$ have sample variances 16 and 25 respectively. Do you accept the hypothesis that $\sigma_1 = \sigma_2$?*

Sol. If $\sigma_1 = \sigma_2$, $[n_1 S_1^2/(n_1-1)]/[n_2 S_2^2/(n_2-1)]$ has an F distribution with n_1-1 and n_2-1 degrees of freedom. Here $n_1=15$, $n_2=10$, $s_1^2=16$ and $s_2^2=25$.

$$\frac{n_1 \; s_1^2/(n_1-1)}{n_2 \; s_2^2/(n_2-1)} = 0.6 \text{ approx.}$$

If $\sigma_1=\sigma_2$, 0.6 is a value assumed by an F with $n_1-1=14$ and $n_2-1=9$ degrees of freedom. The probability of getting an $F_{14,\,9}$ as large as 0.6 is

$$\int_{0.6}^{\infty} f(x)dx$$

where $f(x)$ is the density function of an F with 14 and 19 degrees of freedom. But this probability is greater than 0.05. (This is seen from the tables). If we are ready to accept such a hypothesis, when the probability of getting an F as large as the observed one, is atleast 0.05, we will accept our hypothesis that $\sigma_1=\sigma_2$. The acceptance or rejection depends on the acceptance probability level. This aspect will be discussed in the chapter on Testing Statistical Hypotheses.

Some of the important sampling distributions when a random sample $X_1, X_2, \ldots, X_n$ is taken from a $N(\mu, \sigma)$, are given in the following table.

Statistic	*Distribution*
1. X_i for any i	$N(\mu, \sigma)$
2. $X_1+X_2+\ldots\ldots+X_n$	$N(n\mu, \sqrt{n}\,\sigma)$
3. $\bar{X}=(X_1+X_2+\ldots\ldots+X_n)/n$	$N(\mu, \sigma/\sqrt{n})$
4. $(X_i-\mu)/\sigma$ for any i	$N(0, 1)$
5. $(\bar{X}-\mu)/(\sigma/\sqrt{n})$	$N(0, 1)$
6. $(\bar{X}-\mu)/(S'/\sqrt{n})$ where $S'^2=\sum_{i=1}^{n}(X_i-\bar{X})^2/(n-1)$	Student t with $n-1$ degrees of freedom.
7. $(X_i-\mu)^2/\sigma^2$ for any i	Gamma with parameters $\alpha=1/2$ and $\beta=2$.
8. $\sum_{i=1}^{n}(X_i-\mu)^2/\sigma^2$	Gamma with parameters $\alpha=n/2$ and $\beta=2$. or a χ^2 with n degrees of freedom.
9. nS^2/σ^2 where $S^2=\sum_{i=1}^{n}(X_i-\bar{X})^2/n$	χ^2 with $n-1$ degrees of freedom.
10. $\chi^2_{k_1}+\chi^2_{k_2}+\ldots+\chi^2_{k_p}$ where $\chi^2_{k_1}, \chi^2_{k_2} \ldots \chi^2_{k_p}$ are independent χ^2's	χ^2 with $k_1+k_2+\ldots+k_p$ degrees of freedom.
11. $\dfrac{\chi^2_m/m}{\chi^2_n/n}$ where χ^2_m and χ^2_n are independent χ^2's with m and n d,f., respectively.	F-with m and n degrees of freedom.
12. S'^2_1/S'^2_2 where $S'^2_1=\sum_{i=1}^{n_1}(X_i-\bar{X})^2/(n_1-1)$, $S'^2_2=\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2/(n_2-1)$	F with n_1-1 and n_2-1 d.f.

Where $X_1\ldots, X_{n_1}$ and $Y_1\ldots, Y_{n_2}$ are two independent samples from $N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$ respectively.

Exercises

7.25. Obtain E (F) and Var (F) where F is an F statistic with m and n degrees of freedom.

7.26. If $\int_{F_{\alpha, m, n}}^{\infty} f(x)\, dx = \alpha$, where $f(x)$ is the density function of an F statistic with m and n degrees of freedom, show that

$$F_{\alpha, m, n} = 1/F_{1-\alpha, n, m}.$$

Prove that $F_{m, n}$ and $1/F_{n, m}$ have the same distribution.

7.27. If two random samples of sizes 15 and 20 are taken from a $N(\mu, \sigma)$ what is the probability that the ratio of the sample variances does not exceed 2 ?

7.28. If X and Y are independent show that aX and bY where a and b are non-zero constants and X and Y are s.v's, are also independent.

7.29. If $X_1, \ldots, X_k$ are independent normal variates show that $a_1X_1 + \ldots + a_kX_k$ is normal where a_i 's are constants.

[**Hint** : Use moment generating functions.]

7.30. If $X_1, \cdots, X_k$ are independent normal variates with unit variance and if $\Sigma\, a_ib_j = 0$, show that $a_1X_1 + \cdots + a_kX_k$ and $b_1X_1 + \ldots + b_kX_k$ are independent, where a's and b's are constants.

[**Hint** : If X and Y are normal Cov (X, Y) = 0 implies independence]. Independence of linear combinations of independent variates is a characteristic property of the normal distribution. It can be proved that if two linear forms $Y_1 = a_1X_1 + \ldots + a_kX_k$ and $Y_2 = b_1X_1 + \ldots + b_kX_k$ where X's are independent s.v's, are independent then X_i for which $a_ib_i \neq 0$ is normally distributed. For characterizations of normal distributions by properties like this see reference 3 at the end of this chapter.

7.31. **Variance Stabilizing Transformation.** Let T be a statistic constructed from a sample of size n. Let $E(T) = \theta$ and Var $(T) = \phi(\theta)$. A transformation $T \to g(T)$ or a construction of a function $g(T)$ of T, such that Var $g(T)$ is independent of θ, is called a variance stabilizing transformation. Sometimes such a transformation makes $g(T)$ a normal variate. Under some conditions on T and $g(T)$ it can be thown that such a transformation is given by

$$g(\theta) = \int c\, d\theta / \sqrt{\phi(\theta)}$$

where c is a constant (independent of θ).

(*a*) **Square root transformation.** If X is a Poisson variate with parameter λ show that a variance stabilizing transformation is given by $g(X) = \sqrt{X}$.

(*b*) **Inverse sine transformation.** If X is a Binomial variate with parameters p and n, show that a variance stabilizing transformation for the Binomial proportion X/n is given by $g(X/n) = \sin^{-1} \sqrt{X/n}$.

(*c*) **Tanh^{-1} transformation.** The variance of a sample correlation coefficient r can be shown to be $(1-\rho^2)^2/n$ where n is sufficiently large, and ρ is the population correlation coefficient. Show that a variance stabilizing transformation for r is given by $g(r) = \tanh^{-1} r - (1/2) \log (1+r)/(1-r)$. This is also known as the Z transformation where $z = (1/2) \log (1+r)(1-r)$.

7.32. **Independence of mean and variance in normal samples.** Let $X_1, \ldots, X_n$ be a simple random sample of size n from a $N(\mu, \sigma)$.

(*a*) Write down the joint characteristic function $\phi\ (t_1, t_2)$ of $\overline{X}$ and $S^2 = \Sigma(X_i - \overline{X})^2/n$.

(*b*) For an orthogonal transformation (that is a transformation $XA=Y$ where $AA'=I$, X and Y are vectors and A is a matrix see chapter 1) show that the Jacobian of the transformation is unity in absolute value.

(*c*) Show that there exists an orthogonal transformation of $Y=(x_1, \ldots, x_n)$ into $Y=(y_1, \ldots, y_n)$ such that

$$\Sigma_{j=1}^{n} (x_j-\mu)^2=(y_1-\mu\sqrt{n})^2+\Sigma_{j=2}^{n} y_j^2$$

and

$$\Sigma_{j=1}^{n} (x_j-\bar{x})^2/n = \Sigma_{j=2}^{n} y_j^2/n.$$

(*d*) Now evaluate $\phi(t_1\, t_2)$ by using (*a*), (*b*) and (*c*) and show that it can be put in the form $\phi(t_1, t_2)=\phi_1(t_1)$ and $\phi_2(t_2)$ where ϕ_1 does not contain t_2 and ϕ_2 does not contain t_1.

(*e*) From (*d*) show that $\bar{X}$ and S^2 are independently distributed.

(*f*) From (*e*) obtain the characteristic function of nS^2/σ^2.

(*g*) Show that $\bar{X}$ has a Normal distribution and nS^2/σ^2 has a chi-square distribution with $n-1$ degrees of freedom.

BIBLIOGRAPHY

1. H. **Cramer**, The Mathematical Methods of Statistics. Princeton University Press, 1961.

2. M.G. **Kendall** & A. **Stuart**, The Advanced Theory of Statistics. Vol. I. Charles Griffin and Co., London, 1958.

3. E. **Lukacs** & R.G. **Laha**, Applications of Characteristic Functions Charles Griffin & Co., London, 1964.

4. A.M. **Mood** & F.A. **Graybill**, Introduction to the Theory of Statistics. McGraw-Hill Book Co., New York 1963.

5. C.R. **Rao**, Advanced Statistical Methods in Biometric Research. John Wiley, New York, 1952.

6. C.R. **Rao**, Linear Statistical Inference and Its Applications. Wiley N.Y., 1965.

CHAPTER 8

INTERVAL ESTIMATION

8.0. Introduction.—In the previous chapters we discussed theoretical distributions, sampling and sampling distributions. In the following chapters we will consider the applications of the results derived so far. Statistical methods help us in making a decision in a situation where there is a lack of certainty. This procedure of decision making is usually called statistical inference. Statistical inference may be broadly classified into testing hypotheses and estimation. In this chapter we will consider a special case of estimation problems.

The principle of estimation is to find out estimates for the parameters of a distribution, based on an observed sample from the distribution under consideration. If we give a single quantity as an estimate of a parameter, such an estimate is called a point estimate and the corresponding estimation procedure is called point estimation. This will be discussed in the next chapter. If we estimate an interval such that the interval will cover the true value of the parameter with a certain probability, such an estimation procedure is called interval estimation.

There are many practical situations where we are interested in getting an interval estimate for a parameter. A drug manufacturer may be interested in finding out two numbers so that he can make a statement that, by this new drug the survival rate will be between 90 and 96%. A toothpaste manufacturer would like to estimate the reduction in cavities so that he can claim that his toothpaste will reduce cavities by 40 to 45%. It is helpful for the Incometax Department to have an estimate of the range in which the tax return of the succeeding year will lie. There are many such situations where we would like to get interval estimates.

8.1. CONFIDENCE INTERVALS

If we give an interval estimate for the average I.Q. of all university students on the north American continent and if we say that the average I.Q. is between 105 and 110, we may be making such a statement by observing a random sample of university students. So we can make only a probability statement,

like, the true average I.Q. is between 105 and 110 with a probability of 0.99 etc. This means that our statement will be true in 99% of the cases in the long run or if we go on taking samples of the same size and obtain such intervals, 99% of such intervals will cover the true average I.Q., in the long run. In this case we made a statement with a confidence of 99% and gave an interval (105, 110) which may be called a 99% confidence interval. Such an interval can be given if we know the distribution of the I.Q's.

In general if we find out two quantities t_0 and t_1 based on a random sample from a population $f(x, \theta)$ such that,

$$P\{t_0 \leqslant \theta \leqslant t_1\} = 1 - \alpha \tag{8·1}$$

we say that (t_0, t_1) is a $100(1-\alpha)\%$ confidence interval for θ; t_0 and t_1 are called the lower confidence limit and the upper confidence limit respectively and $(1-\alpha)$ is called the confidence coefficient. Such a statement can be made about a parameter by suitably selecting a statistic which contains the parameter under consideration and whose distribution is independent of the parameters. This can be seen from the following examples. But such a statistic need not exist always. Further in this chapter we will construct confidence intervals only for the parameters in a normal distribution and in a Binomial distribution. The same ideas can be used for setting up confidence intervals for the parameters of other distributions also.

Ex. 8.1.1. *An observed random sample of size 9 from a $N(\mu, \sigma=2)$ has a mean 50, obtain a 95% confidence interval for μ.*

Sol. Here we are asked to make a statement,

$$P\{t_0 \leqslant \mu \leqslant t_1\} = 0·95,$$

where t_0 and t_1 are known quantities.

Let us consider the statistic

$$t = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{8.2}$$

This contains the parameter μ. We have an observed value of $\overline{X}$ and σ and n are known. Further t is a $N(0, 1)$ and hence its distribution is independent of the parameters. From a normal table we get 1.96 such that

$$P\left\{-1.96 \leqslant \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leqslant 1.96\right\} = 0.95 \tag{8.3}$$

This is illustrated in Fig. 8.1.

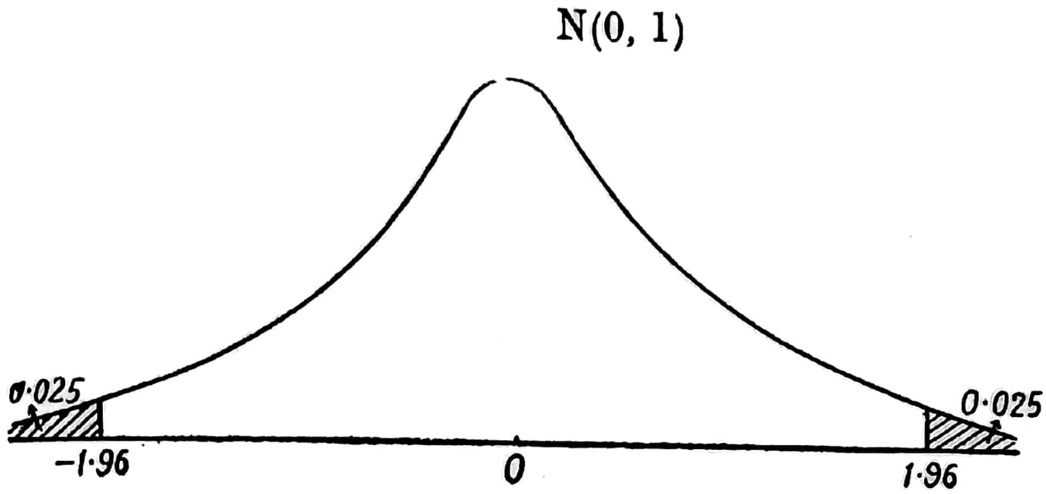


Fig. 8.1.

The double inequality

$$-1.96 \leqslant \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leqslant 1.96 \tag{8.4}$$

may be written as

$$-1.96 \cdot \frac{\sigma}{\sqrt{n}} \leqslant \bar{x}-\mu \leqslant 1.96 \frac{\sigma}{\sqrt{n}}$$

i.e.,

$$\bar{x}-1 \cdot 96 \frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{x}+1.96 \frac{\sigma}{\sqrt{n}} \tag{8.5}$$

But in this problem $\bar{x}=50$, $\sigma=2$ and $n=9$ and therefore

$$48.693 \leqslant \mu \leqslant 51.307$$

or

$$P\{48.693 \leqslant \mu \leqslant 51.307\}=0.95 \tag{8.6}$$

48.693 and 51.307 are the lower and upper 95% confidence limits and (48.693, 51.307) is a 95% confidence interval for μ.

Comments. The probability statement (8.6) does not mean that μ is a stochastic variable and it will fall in the interval (48.693, 51.307), it means that if we continue taking random samples of size 9 and every time calculating a 95% confidence interval for μ, 95% of our intervals will cover μ, in the long run. Here μ is an unknown constant ; we find an interval such that most probably this unknown value is on this interval.

8.2. THE BEST CONFIDENCE INTERVAL

In Ex. 8.1.1 we obtained a 95% confidence interval for μ as (48.693, 51.307). This is obtained by deleting areas equal to 0·025 at both tails of a standard normal distribution (Fig. 8.1) It can be seen that the interval given above is not a unique one. Suppose that we had deleted an area equal to 0.05 at the right end then we can get two quantities $-\infty$ and 1.64 from a normal table such that

$$P\left\{-\infty < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leqslant 1.64\right\}=0.95 \tag{8·7}$$

This leads to the interval $(-\infty, \bar{x}+1.64\ \sigma/\sqrt{n})$

$$=(-\infty, 51.09) \tag{8.8}$$

A number of 95% confidence intervals can be constructed for μ. Evidently we would prefer to have the one which is shortest. This is one criterion of comparison of intervals constructed with the same confidence coefficient. For other desirable properties of confidence intervals, such as 'short on the average', 'most selective' 'short unbiased' etc., the reader may refer to the bibliography at the end of this chapter. In a symmetrical distribution it may be easily seen that the central intervals, in the sense the intervals obtained by deleting equal areas at both ends, usually give the shortest intervals. This is illustrated by Fig. 8.2.

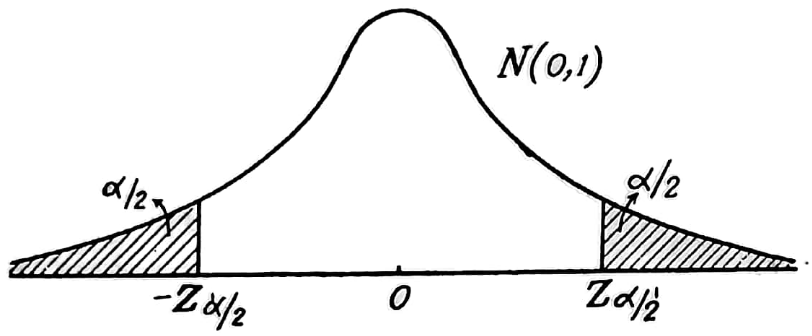


Fig 8.2.

In this illustration if $Z_{\alpha/2}$ is moved to the left so as to cut off an area equal to $(2/3)\alpha$ at the right tail then $-Z_{\alpha/2}$ is moved more to the left so as to make the sum of the two tail areas equal to α. This results in a longer interval than the interval corresponding to the omission of equal tail areas at both ends. In the following sections only the central intervals are considered. Here and in the following sections we consider the construction of confidence intervals when there is only one parameter. The case when there is more than one parameter is discussed in section 8.8.

8.3. CONFIDENCE INTERVALS FOR MEANS

Let $x_1, x_2, \ldots x_n$ be an observed random sample from a $N(\mu, \sigma)$ where σ is known. A $100(1-\alpha)\%$ confidence interval may be established for μ based on the observed sample.

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} : N(0, 1) \tag{8.9}$$

Corresponding to any α we can find out an $Z_{\alpha/2}$ such that

$$P\left\{-Z_{\alpha/2} \leqslant \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leqslant Z_{\alpha/2}\right\} = 1-\alpha.$$

This is illustrated in Fig. 8.3.

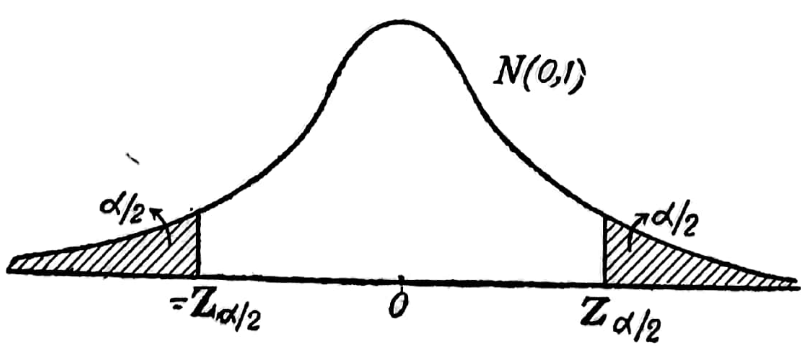


Fig. 8.3.

$Z_{\alpha/2}$ is obtained from a normal probability table.

$$P\left\{-Z_{\alpha/2} \leqslant \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leqslant Z_{\alpha/2}\right\}=1-\alpha \tag{8.10}$$

$$P\left\{-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leqslant \bar{x}-\mu \leqslant Z_{\alpha/2}\frac{\sigma}{\sqrt{\ }}\right\}=1-\alpha$$

$$P\left\{\bar{x}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{x}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}=1-\alpha \tag{8.11}$$

Here σ is known and hence $100(1-\alpha)\%$ confidence limits $\bar{x}-Z_{\alpha/2}\,\sigma/\sqrt{n}$ and $\bar{x}+Z_{\alpha/2}\,\sigma/\sqrt{n}$, for μ, are known. This $100(1-\alpha)\%$ confidence interval for μ is

$$(\bar{x}-Z_{\alpha/2}\,\sigma/\sqrt{n},\ \bar{x}+Z_{\alpha/2}\,\sigma/\sqrt{n}) \tag{8.12}$$

If σ is not known then the statistic

$$t_{n-1}=\frac{(\bar{X}-\mu)}{S'/\sqrt{n}} \tag{8.13}$$

has a student t distribution with $n-1$ degrees of freedom, where $S'^2=\Sigma\,(X_i-\bar{X})^2/(n-1)$. Hence corresponding to any α we can find out a $t_{\alpha/2}$ such that

$$P\left\{-t_{\alpha/2} \leqslant \frac{\bar{x}-\mu}{s'/\sqrt{n}} \leqslant t_{\alpha/2}\right\}=1-\alpha. \tag{8.14}$$

This is illustrated in Fig. 8.4.

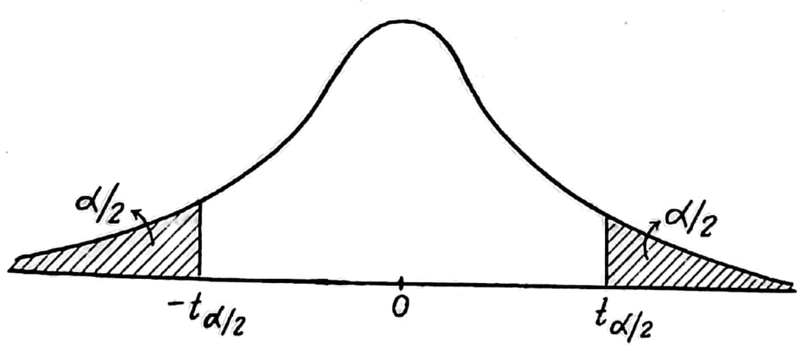


Fig. 8.4.

Fig. 8.4 gives the distribution of a student t with $n-1$ degrees of freedom

$$P\left\{-t_{\alpha/2} \leqslant \frac{\bar{x}-\mu}{s'/\sqrt{n}} \leqslant t_{\alpha/2}\right\}=1-\alpha.$$

i.e., $\quad P\{-t_{\alpha/2}\,s'/\sqrt{n} \leqslant \bar{x}-\mu \leqslant t_{\alpha/2}\,s'/\sqrt{n}\}=1-\alpha$

$$\therefore \quad P\{\bar{x}-t_{\alpha/2}\,s'/\sqrt{n} \leqslant \mu \leqslant \bar{x}+t_{\alpha/2}\,s'/\sqrt{n}\}=1-\alpha. \tag{8.15}$$

The $100(1-\alpha)\%$ confidence interval for μ is

$$(\bar{x}-t_{\alpha/2}\,s'/\sqrt{n},\ \bar{x}+t_{\alpha/-}\,s'/\sqrt{n})$$

where $t_{\alpha/2}$ is obtained from a student t table.

Ex. 8.3.1. *A random sample of 25 experimental beef cattle showed an average increase of 35 lbs after administering a new diet. If the experimenter has enough data to justify the assumption that the increase in weight is approximately normal with standard deviation 4, obtain a 95% confidence interval for the expected increase in weight by the new diet.*

Sol. According to our notation, the confidence coefficient, $1-\alpha=0.95$, which implies that $\alpha=0.05$ or $\alpha/2=0.025$.

Consider the statistic,

$$t=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \tag{8.16}$$

But t is a N(0, 1), and thus the distribution of t is independent of the parameter μ,

$$f(t)=(2\pi)^{-\frac{1}{2}}\, e^{-t^2/2},\ -\infty<t<\infty \tag{8.17}$$

$$\text{But, } \int_{1.96}^{\infty} f(t)\, dt=0.025=\int_{-\infty}^{-1.96} f(t)\, dt. \tag{8·18}$$

1.96 is obtained from a normal probability table. Therefore,

$$P\left\{-1.96\leqslant \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leqslant 1.96\right\}=0.95. \tag{8.19}$$

That is, $P\{\bar{x}-1.96\ \sigma/\sqrt{n} \leqslant\mu\leqslant\bar{x}+1.96\ \sigma/\sqrt{n}\}=0.95.$ (8.20)

Here $\bar{x}=35$, $\sigma=4$, and $n=25$.

Therefore a 95% confidence interval for μ is,

$$(\bar{x}-1{\cdot}96\ \sigma/\sqrt{n},\ \bar{x}+1.96\ \sigma/\sqrt{n})=(33.43,\ 36.57) \tag{8.21}$$

Comments. According to our results, $P\{33.43\leqslant\mu\leqslant 36.57\}=0.95$ does not mean that μ is a stochastic variable and the probability that it will fall between 33.43 and 36.57 is 0.95. The meaning of our confidence statement is as follows; if sampling is continued, each time taking a random sample of size 25, we can evaluate the interval $(\bar{x}-1.96\ \sigma/\sqrt{n},\ \bar{x}+1.96\ \sigma/\sqrt{n})$ for every sample. In the long run 95% of the intervals will cover the true value of the parameter μ or 95% of the intervals will contain μ. If we make a statement that μ lies in one of these intervals we will be wrong in 5% of the cases, in the long run.

Ex. 8.3.2. *A dress-maker finds that a random sample of 16 girls of a certain age group in a particular city shows an average bust measurement of 40" with a variance of 25". If he has enough evidence to assume that the bust measurements are normally distributed obtain a 99% interval estimate for the average bust measurement of girls in that age group in that city.*

Sol. According to our notation $n=16$, $\bar{x}=40$ and $s^2=25$. In this problem the population variance σ^2 is not known and hence we will consider the statistic,

$$t=\frac{\bar{X}-\mu}{S'/\sqrt{n}} \tag{8.22}$$

where $S'^2=\Sigma(X_i-\bar{X})^2/(n-1)$ and t is a student t with $n-1$ degrees of freedom and further the distribution of t is independent of μ and the 'nuisance' parameter σ.

$$1-\alpha=0.99, \quad \Rightarrow \quad \alpha/2=0.005.$$

From a student t table corresponding $n-1=16-1=15$ degrees of freedom, we get the values -2.947 and 2.947 such that

$$\int_{-\infty}^{-2.947} f(t)\,dt=0.005=\int_{2.947}^{\infty} f(t)\,dt \tag{8.23}$$

where $f(t)$ is the density function of a student t with 15 degrees of freedom.

$$\text{Hence} \quad P\left\{-2.947\leqslant \frac{\bar{x}-\mu}{s'/\sqrt{n}} \leqslant 2.947\right\}=0.99. \tag{8.24}$$

i.e.

$$P\left\{\bar{x}-2.947.\frac{s'}{\sqrt{n}} \leqslant \mu\leqslant\bar{x}+2.947\frac{s'}{\sqrt{n}}\right\}=0.99. \tag{8.25}$$

Here $\bar{x}=40$, $n=16$ and $s^2=25=\Sigma(x_i-\bar{x})^2/n$.

But $ns^2/(n-1)=s'^2=(16)(25)/15=26.67$

A 99% confidence interval for μ is

$$\left(\bar{x}-2.947\frac{s'}{\sqrt{n}},\ \bar{x}+2.947\frac{s'}{\sqrt{n}}\right)=\left(36,96,\ 43.04\right) \tag{8·26}$$

Comments. Selection of the appropriate statistic is the most important problem. The statistic should contain the parameter under consideration and the distribution of the statistic should be independent of the parameter. It may be noticed that the 99% confidence interval is not unique. From a student t table, corresponding to 15 degrees of freedom, we can find out quantity 2.602 such that

$$\int_{2.602}^{\infty} f(t)\,dt=0.01 \Rightarrow \int_{-\infty}^{2.602} f(t)\,dt=0.99$$

i.e.,

$$P\left\{-\infty\leqslant \frac{\bar{x}-\mu}{s'/\sqrt{n}} \leqslant 2.602\right\}=0.99$$

i.e., $$P\left\{ -\infty \leqslant \mu \leqslant \bar{x} + 2.502 \frac{s'}{\sqrt{n}} \right\} = 0.99.$$

Here a 99% confidence interval for μ is obtained as

$$\left(-\infty, \bar{x} + 2.602 \frac{s'}{\sqrt{n}} \right) = \left(-\infty, 42.68 \right). \tag{8.27}$$

Evidently this interval is larger than the one we obtained before. Central intervals obtained from equations of the type

$$P\{-t_{\alpha/2} \leqslant T \leqslant t_{\alpha/2}\} = 1 - \alpha,$$

are usually the shortest in symmetric distributions, where T is the statistic under consideration.

Exercises

8.1. A random sample of size 50, taken from a $N(\mu, \sigma = 5)$, has a mean 40. Obtain a 95% confidence interval for $2\mu + 3$.

8.2. The average height of 20 university students is 65″. Assuming these observations as a random sample from a $N(\mu, \sigma = 2)$ obtain a 90% interval estimate of the average height of all the university students from which the sample is taken.

8.3. A random sample of 25 citizens in a country shows an average annual income of \$ 10,000 with a standard deviation of \$ 200. Assuming the income distribution is a $N(\mu, \sigma)$, obtain a 90% interval estimate for the average income of the citizens in the country.

8.4. Ten bullets from an enemy gun have an average diameter of 5 units with a standard deviation of 0.02 units. Assuming that this sample is a random sample from a $N(\mu, \sigma)$, obtain a 99% interval estimate of the diameter of the enemy gun barrel, taking the diameter of the gun barrel $= \mu + 0.01$. Is it possible to get an interval estimate if (1) only one bullet is available, (2) only one bullet is available, but σ is known to be 0.03.

8.5. A random sample of 40 chickens from a farm, has an average weight of 4 lbs with a standard deviation of 0.5 lb. Assuming that the sample can be considered to be from a $N(\mu, \sigma)$, give a 99% interval estimate of the expected income of the farmer on the average per chicken if chickens are sold \$0.25 a lb.

8.6. A random sample of 35 days show an average increase of \$50 with a standard deviation of \$10 in sales after the appointment of a new sales girl. Assuming that the increase in sales can be considered to be distributed as a $N(\mu, \sigma)$, obtain a 95% interval estimate for the expected increase in sales per day after the new appointment.

8.7. A survey conducted on a particular day among a random sample of 20 students from a particular university shows that they have spent on the average 5.5 hours for studies with a standard deviation of 0.5 hour. Assuming that the number of hours spent on that day by the students of that university, has a $N(\mu, \sigma)$, obtain a 95% confidence interval for the expected decrease in the expenses, taking the decrease in expenses = (0.1 time the number of hours spent for studies on that day).

8.4. CONFIDENCE INTERVALS FOR DIFFERENCE BETWEEN MEANS

Let $x_1, x_2, \ldots x_{n_1}$ and $y_1, \ldots, y_{n_2}$ be two independent random samples from the normal populations $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$

respectively, where σ_1 and σ_2 are known. Based on these observed samples we can construct $100(1-\alpha)\%$ confidence intervals for $\mu_1-\mu_2$.

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^{\frac{1}{2}}} : N(0,\ 1) \tag{8.28}$$

Hence corresponding to any α we can find out an $Z_{\alpha/2}$ from normal tables such that

$$P\{-Z_{\alpha/2} \leqslant \frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^{\frac{1}{2}}} \leqslant Z_{\alpha/2}\}=1-\alpha. \tag{8.29}$$

$$P\left((\bar{x}-\bar{y})-Z_{\alpha/2}\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^{\frac{1}{2}} \leqslant \mu_1-\mu_2 \leqslant (\bar{x}-\bar{y}) + Z_{\alpha/2}\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^{\frac{1}{2}}\right)=1-\alpha \tag{8.30}$$

Here σ_1 and σ_2 are known, then a $100(1-\alpha)\%$ confidence interval for $\mu_1-\mu_2$ is

$$\left(\bar{x}-\bar{y}-Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\quad \bar{x}-\bar{y}+Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_1}}\right) \tag{8.31}$$

If $\sigma_1=\sigma_2=\sigma$ and if σ is unknown then

$S^2=\left(n_1 S_1^2+n_2S_2^2\right)/\left(n_1+n_2-2\right)$ is an unbiased estimator for σ^2 and hence,

$$t_{n_1+n_2-2}=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S/\sqrt{(1/n_2)+(1/n_2)}} \tag{8.32}$$

is a student t with n_1+n_2-2 degrees of freedom and if n_1+n_2-2 is sufficiently large $t_{n_1+n_2-2}$ has an approximate standard normal

distribution, where $S_1^2 = \Sigma_{i=1}^{n_1} (X_i - \bar{X})^2/n_1$

and $S_2^2 = \Sigma_{i=1}^{n_2} (Y_i - \bar{Y})/n_2$. Therefore a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}-\bar{y}-t_{\alpha/2}\ s/\ \sqrt{1/n_1+1/n_2}\ ,\ \bar{x}-\bar{y}+t_{\alpha/2}\ s/\ \sqrt{1/n_1+1/n_2}) \tag{8.33}$$

where $t_{\alpha/2}$ is obtained from a student t table corresponding to n_1+n_2-2 degrees of freedom. If n_1+n_2-2 is sufficiently large a $100(1-\alpha)\%$ confidence interval for $(\mu_1-\mu_2)$ is given by

$$(\bar{x}-\bar{y} \pm z_{\alpha/2}\ s/\sqrt{1/n_1+1/n_2}) \tag{8.34}$$

where $z_{\alpha/2}$ is obtained from a normal table. (A good approximation to the normal is obatined when $n_1+n_2-2 > 30$).

Ex. 8.4.1. *A farmer made the following observations. Random samples of 10 and 12 newly planted rubber plants of two varieties gave the average growths in the first week as 25" and 24" respectively. If he has evidence to assume that the growths are distributed as $N(\mu_1, \sigma_1=2)$ and $N(\mu_2, \sigma_2=3)$ respectively, obtain a 95% confidence interval for the expected growth difference in the first week.*

Sol. According to our notation $n_1=10$, $n_2=12$, $\sigma_1=2$, $\sigma_2=3$ $\bar{x}_1=25$ and $\bar{x}_2=24$.

$$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\left(\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}\right)^{1/2}} : N(0,1)$$

Therefore from the normal tables we get a quantity 1.96 such that

$$P\left\{-1.96 \leqslant \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\left(\dfrac{\sigma_1^2}{n_1}\ \ \dfrac{\sigma_2^2}{n_2}\right)^{1/2}} \leqslant 1.96\right\} = 0.95\text{-}$$

i.e., $$P\left\{(\bar{x}_1-\bar{x}_2)-1.96\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^{1/2} \leqslant \mu_1-\mu_2 \leqslant (\bar{x}_1-\bar{x}_2)+1.96\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^{1/2}\right\}=0.95.$$

A 95% confidence interval for $\mu_1-\mu_2$ is,

$$\left(\bar{x}_1-\bar{x}_2-1.96\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\,,\ \bar{x}_1-\bar{x}_2+1.96\sqrt{\frac{\sigma_1^2}{n_1}\quad\frac{\sigma_2^2}{n_2}}\right)$$

$$=(-1.097,\ 3.097) \tag{8.35}$$

Comments. It may be noticed that this 95% confidence interval is not a unique one. We could have found two values t_0 and t_1 such that

$$P\left\{t_0 \leqslant \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\left(\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}\right)^{1/2}} \leqslant t_1\right\}=0.95 \tag{8.36}$$

This would have led to a different interval estimate if t_0 and t_1 were different from -1.96 respectively. If σ_1 and σ_2 are not given these 'nuisance' parameters can be avoided by taking the sample variances as estimates for σ_1^2 and σ_2^2 when the sample sizes are large.

Ex. 8.4.2. *Two independent random samples of sizes 10 and 12 from the populations $N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$, have means 50 and 45 and variance 16 and 25 respectively. Obtain a 90% confidence interval for $\mu_1-\mu_2$.*

Sol. According to our notation

$$n_1=10,\ n_2=12,\ \bar{x}_1=50,\ \bar{x}_2=45,$$

$$s_1^2=16,\ s_2^2=25,\ \frac{\alpha}{2}=0.05.$$

$$E(\bar{X}_1-\bar{X}_2)=\mu_1-\mu_2$$

and

$$\text{Var}\,(\bar{X}_1-\bar{X}_2)=\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$

and σ^2 may be estimated by

$$S^2=\left(n_1\,S_1^2+n_2\,S_2^2\right)\big/(n_1+n_2-2)$$

Hence

$$t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{S^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}} \tag{8.37}$$

is a student t statistic with $n_1+n_2-2=20$ degrees of freedom. From a student t table, corresponding to 20 degrees of freedom we get a value 1.725, such that

$$\int_{-\infty}^{-1.725} f(t)dt = 0.05 = \int_{1.725}^{\infty} f(t)dt \tag{8.38}$$

where $f(t)$ is the density function of a student t with 20 degrees of freedom. Therefore

$$P\left\{-1.725 \leqslant \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \leqslant 1.725\right\} = 0.90 \tag{8.39}$$

$$P\left((\bar{x}_1-\bar{x}_2)-1.725\,.\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \leqslant \mu_1-\mu_2 \leqslant (\bar{x}_1-\bar{x}_2)\right.$$

$$\left.+1.725\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right\} = 0.90$$

i.e.,

$$P\{3.483 \leqslant \mu_1-\mu_2 \leqslant 6.517\} = 0.90 \tag{8.40}$$

Hence a 90% confidence interval for $\mu_1-\mu_2$ is (3.483, 6.517).

Comments. If $n_1+n_2-2>30$ instead of a student t we may use a normal approximation. If the population variances are different for large sample sizes of n_1 and n_2, the statistic

$$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^{1/2}} \tag{8.41}$$

is approximately a N(0, 1). Hence in such a case this approximation may be used to set up confidence intervals.

Exercises

8.8. Independent random samples of sizes 20 and 25 taken from $N(\mu_1, \sigma_1=3)$ and $N(\mu_2, \sigma_2=4)$ have means 50 and 45 respectively. Construct a 90% confidence interval for $2(\mu_1-\mu_2)$.

8.9. Independent random samples of 10 girls and 12 boys of a certain age group have average I·Q's 104 and 103, with standard deviations 1.1 and 1.2 respectively. Assuming that the I.Q's are distributed as $N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$ respectively, obtain a 95% interval estimate for the expected difference in the I.Q's of girls and boys.

8.10. A survey conducted on random samples of sizes 50 and 70 people of a certain profession in two cities, shows the average incomes as \$100 and \$80 per week with standard deviations \$5 and \$2 respectively. Assuming that the incomes of the people of this profession are distributed as $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$ respectively, obtain a 95% interval for the expected difference of incomes for this profession in these cities.

[**Hint.** Use a normal approximation].

8.11. A random sample of 40 days shows the average output of a production process as 30 units by method A and 35 units by method B with standard deviations 5 units and 7 units respectively. Under the assumption of normality obtain a 99% interval estimate for the expected difference in the outputs by the two methods.

8.5. CONFIDENCE INTERVALS FOR PROPORTIONS

Let us consider a binomial probability situation. Let p be the probability of a success in any trial and let x be the number of successes observed in N trials. Then $\hat{p}=x/N$ is the observed proportion of successes.

But $$E(\hat{p})=E\left(\frac{X}{N}\right)=p \text{ and } \operatorname{Var}(\hat{p})=p(1-p)/N \tag{8.42}$$

For large N, $\dfrac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{N}}}$ is approximately normally distributed or

$$\frac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{N}}} : N(0, 1) \tag{8.43}$$

Corresponding to any given α, $z_{\alpha/2}$ can be found out from normal tables, such that,

$$P\left\{-z_{\alpha/2}\leqslant \frac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{N}}} \leqslant z_{\alpha/2}\right\}=1-\alpha \tag{8.44}$$

The inequality

$$-z_{\alpha/2}\leqslant \frac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{N}}}\leqslant z_{\alpha/2} \tag{8.45}$$

upon simplification, may be transformed to

$$\frac{x+\frac{1}{2}z^2_{\alpha/2}-z_{\alpha/2}\sqrt{x\dfrac{(N-x)}{N}+\dfrac{1}{4}Z^2_{\alpha/2}}}{N+z^2_{\alpha/2}}$$

$$\leqslant p \leqslant \frac{x+\frac{1}{2}z^2 \quad +z_{\alpha/2}\sqrt{x\dfrac{(N-x)}{N}+\dfrac{1}{4}\quad z_{\alpha/2}}}{N+z^2_{\alpha/2}} \tag{8.46}$$

where $x/N=\hat{p}$. Hence a $100(1-\alpha)\%$ confidence interval for p is given by,

$$\frac{x+\left(\dfrac{1}{2}\right)z^2_{\alpha/2}\pm z_{\alpha/2}\sqrt{x\dfrac{(N-x)}{N}+\left(\dfrac{1}{4}\right)z^2_{\alpha/2}}}{N+z^2_{\alpha/2}} \tag{8.47}$$

When N is sufficiently large $(\hat{p}-p)/\sqrt{\hat{p}(1-\hat{p})/N}$ may be assumed to be approximately N(0, 1). Then

$$P\left\{-z_{\alpha/2} \leqslant \frac{(\hat{p}-p)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}} \leqslant z_{\alpha/2}\right\}=1-\alpha \tag{8.48}$$

That is $P\{\hat{p}-z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/N} \leqslant \hat{p} \leqslant p+z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/N}\}=1-\alpha$

Hence a $100(1-\alpha)\%$ confidence interval for p is,

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/N} \tag{8.49}$$

If the total number of trials N is not large, corresponding to a given α we can find out numbers t_0 and t_1 from binomial probability tables so that,

$$P\{t_0 \leqslant p \leqslant t_1\}=1-\alpha \tag{8.50}$$

It is seen that transformation of equation (8.45) to (8.46) involves the problem of solving quadratic equations. Sometimes such a separation of the patameter under consideration poses greater difficulties. In such situations a graphical representation usually gives some ideas about the confidence intervals.

Ex. 8.5.1. *In a random sample of 100 articles, 10 are found defective. Obtain a 95% confidence interval for the true proportion of defectives in the population of such articles under consideration.*

Sol. The proportion $\hat{p}$ of defectives in the sample

$$=\frac{10}{100}=0.1$$

Here the sample size N is large and hence we can assume that the statistic

$$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}} \text{ is approximately a } N(0, 1).$$

Therefore from normal probability tables, we obtain 1.96 such that

$$P\left\{-1.96 \leqslant \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}} \leqslant 1.96\right\}=0.95$$

i.e. $$P\left\{\hat{p}-1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} \leqslant p \leqslant \hat{p}+1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}\right]=0.95$$

Hence a 95% confidence interval for p is

$$\left(\hat{p}-196\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}\ ,\ \hat{p}+1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}\right)$$

$$=(-0.086,\ 0.286) \tag{8.51}$$

Comments. When N is not large we can obtain two values t_0 and t_1 by using a table of binomial probabilities, such that

$$P\{t_0 \leqslant p \leqslant t_1\}=0.95.$$

This often involves complications since we may not be able to find out a statistic involving p with its distribution independent of p. In such a case we can adopt a general procedure. For additional reading in this line see reference [5] at the end of this chapter.

Exercises

8.12. A random sample of 40 articles of a particular type shows that 5 of them do not meet quality specifications. Obtain a 99% interval estimate for the expected number of defective ones in a shipment of 10,000 such articles.

8.13. A random sample of 100 seagull eggs collected from an island shows that 10 of them are spoiled or will not hatch. If there are 10,000 eggs on that island, obtain a 95% interval estimate for the expected number of chicks.

8.14. Two random samples of sizes 100 each of a particular article from two production processes show that 2% are defective by one process and 3% are defective by the other process. Obtain a 90% interval estimate for the expected difference in the proportion of defective articles by the two processes.

8.15. A survey conducted on two random samples of sizes 100 each shows that the survival rates from a disease, is 90% by drug A and 95% by drug B. Obtain a 99% interval estimate for the expected difference in the survival rates by the two drugs.

8.6. CONFIDENCE INTERVAL FOR VARIANCE

Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu, \sigma)$ we shall construct $100(1-\alpha)\%$ confidence intervals for σ^2.

$$n\, S^2/\sigma^2 \ ;\ \chi^2_{n-1} \tag{8.52}$$

where $\qquad S^2=\Sigma(X_i-\bar{X})^2/n \quad$ and χ^2_{n-1}

is a χ^2 with $n-1$ degrees of freedom. From a χ^2 table we can find out two values $\chi^2_{1-\alpha/2}$ and $\chi^2_{\alpha/2}$ such that

$$P\left\{\chi^2_{1-\alpha/2} \leqslant \sigma^{\frac{nS^2}{\sigma^2}} \leqslant \chi^2_{\alpha/2}\right\}=1-\alpha$$

This is illustrated in Fig. 8.5.

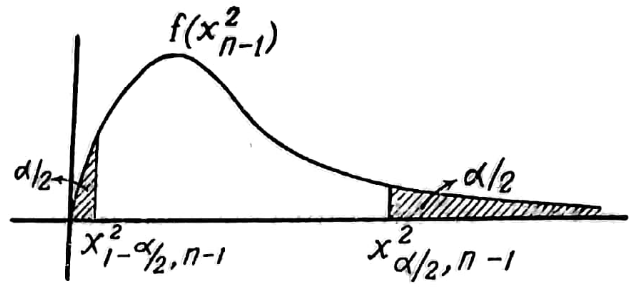


Fig. 8.5.

The double inequality

$\chi^2_{1-\alpha/2} \leqslant \frac{n s^2}{\sigma^2} \leqslant \chi^2_{\alpha/2}$ may be written as

$$\frac{\chi^2_{1-\alpha/2}}{n s^2} \leqslant \frac{1}{\sigma^2} \leqslant \frac{\chi^2_{\alpha/2}}{n s^2} \tag{8.53}$$

i.e.,

$$\frac{ns^2}{\chi^2_{\alpha/2}} \leqslant \sigma^2 \leqslant \frac{ns^2}{\chi^2_{1-\alpha/2}} \tag{8.54}$$

Hence a $100(1-\alpha)\%$ confidence interval for σ^2 is,

$$\left(ns^2/\chi^2_{\alpha/2}\ ,\ ns^2/\chi^2_{1-\alpha/2}\right) \tag{8.55}$$

or a $100(1-\alpha)\%$ confidence interval for σ may be given as,

$$\left(\sqrt{ns^2/\chi^2_{\alpha/2}}\ ,\ \sqrt{ns^2/\chi^2_{1-\alpha/2}}\right) \tag{8.56}$$

Ex. 8.6.1. *A random sample of 20 citizens from a township shows that the variance of their daily incomes is \$36. Assuming that the income distribution there, is approximately a $N(\mu, \sigma)$ obtain a 95% interval estimate for σ.*

Sol. Here $n=20,\ s^2=36,\ 1-\alpha=0.95 \Rightarrow \alpha/2=0.025$

$$nS^2/\sigma^2 : \chi^2_{n-1}$$

From a χ^2 table, corresponding to $n-1=19$ d.f., we get the quantities 8.907 and 32.852, such that,

$$\int_0^{8.907} f\left(\chi^2_{19}\right) d\chi^2_{19} = 0.025 = \int_{32.852}^{\infty} f\left(\chi^2_{19}\right) d\chi^2_{19} \tag{8.57}$$

where $f\left(\chi^2_{19}\right)$ is the density function of a chi-square with 19 d.f.

Therefore,

$$P\{8.907 \leqslant ns^2/\sigma^2 \leqslant 32.852\} = 0.95$$

That is, $P\{ns^2/32.852 \leqslant \sigma^2 \leqslant ns^2/8.907\} = 0.95.$

Hence a 95% confidence interval for σ^2 is,

$$(ns^2/32.852,\ ns^2/8.907) = (21.92.\ 80.83). \qquad (8.58)$$

Comments. From this confidence interval a 95% confidence interval for σ may be obtained as,

$$(\sqrt{21.92}\ ;\ \sqrt{80.83}). \qquad (8.59)$$

It may be noticed that the interval given above is not a unique one. For small sample sizes the χ^2 distribution is not a symmetric one. So there is no guarantee that the central interval is the shortest one.

Exercises

8.16. A random sample of 20 bullets produced by machine shows a standard deviation of 0.2 mm. in the measurement of their diameters. Assuming that the diameter measurement is a $N(\mu, \sigma)$ obtain a 95% interval estimate of σ.

8.17. A random sample of 10 housewives in a city shows an average weight of 135 lbs with a standard deviation of 5 lbs. Assuming normality for the weight measurements obtain a 99% interval estimate for the true variance σ^2.

8.18. A random sample of 12 married teenage girls shows an average I.Q. of 90 with a standard deviation of 2. Assuming that the I.Q's of such girls have a $N(\mu, \sigma)$ obtaln a 99% interval estimate for 2σ.

8.19. The standard deviation of the marks obtained by a random sample of 20 freshmen in a particular university is 10. Assuming that the marks of freshness in this university are approximately a $N(\mu, \sigma)$, obtain a 95% interval for $3\sigma^2$.

8.20. If a random sample of size n is taken from a $N(\mu, \sigma)$ then $S' = [\Sigma(X_i - X)^2/(n-1)]^{1/2}$ is approximately normally distributed with mean σ and with variance $\sigma^2/2n$. Construct a $100(1-\alpha)\%$ confidence interval for σ by using this approximation.

8.7. SUMMARY

The following table on page 269 gives some of the confidence intervals for the parameters mentioned in the same table. These intervals are based on an observed random sample of size n in the case of a single population, and two independent random samples of sizes n_1 and n_2 in the case of two populations,

$$Z_{\alpha/2},\ t_{\alpha/2,\,v},\ \chi^2_{\alpha/2,\,v},\ \chi^2_{1-\alpha/2,\,v}$$

S, S′, S″, are defined by,

$$\int_{Z\alpha/2}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \alpha/2, \quad \int_{t\alpha/2,\, v}^{\infty} f(t)dt = \alpha/2$$

$$\int_{0}^{\chi^2_{1-\alpha/2,\, v}} g(x)dx = \alpha/2 = \int_{\chi^2_{\alpha/2}}^{\infty} g(x)\, dx$$

$$S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/n, \; S'^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1)$$

$$S''^2 = \left(n_1 S_1^2 + n_2 S_2^2 \right) \Big/ (n_1 + n_2 - 2),$$

$$S_1^2 = \sum_{i=1}^{n_1} (X_i - \bar{X})^2/n_1, \; S_2^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2/n_2$$

respectively, where $f(t)$ and $g(x)$ are the density functions of a student t with v degrees of freedom and a χ^2 with v degrees of freedom respectively.

8.8. CONFIDENCE REGIONS

In the previous sections we have seen that we can establish confidence intervals for a parameter in a probability distribution, if we can get a statistic whose distribution is independent of the parameter. In other words we can estimate a parameter in terms of an interval such that the interval will contain the true value of the parameter, with a probability of $1-\alpha$, for any given α. In this section we will consider the interval estimation of the parameters of a probability distribution when there is more than one parameter For example in a $N(\mu, \sigma)$ there are two parameters μ and σ. Based on an observed sample, if a region is constructed such that the region will cover the true diameter values, with a probability $1-\alpha$, then we can say that the proposed region is a $100(1-\alpha)\%$ confidence region for the parameters μ and σ. Here also our success in constructing a confidence region depends on the selection of a suitable statistic which contains the parameters under consideration, but whose distribution is independent of the 100 parameters. The idea of a $(1-\alpha)\%$ confidence region is illustrated in Fig. 8.6. where C is a region in the parameter space (that is, the space generated by all possible values of the parameters. In our example $-\infty<\mu<\infty$, $0<\sigma<\infty$ and hence the parameter space, is the upper half plane. If there are three parameters

For the Parameter	*Statistic*	$100(1-\alpha)$ % *confidence interval*
1. μ in $N(\mu, \sigma)$ (σ is known)	$(\overline{X}-\mu)/(\sigma/\sqrt{n})$	$\overline{x} \pm Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$
2. μ in $N(\mu, \sigma)$ (σ unknown sample size large)	$(\overline{X}-\mu)/(S/\sqrt{n})$	$\overline{x} \pm Z_{\alpha/2} \dfrac{s}{\sqrt{n}}$
3. μ in $N(\mu, \sigma)$ (σ unknown, sample size small)	$(\overline{X}-\mu)/(S'/\sqrt{n})$	$\overline{x} \pm t_{\alpha/2,\, n-1} \dfrac{s'}{\sqrt{n}}$
4. $\mu_1-\mu_2$ in $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$ (σ_1 and σ_2 known)	$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\left(\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}\right)^{1/2}}$	$(\overline{x}-\overline{y})+Z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$
5. $\mu_1-\mu_2$ in $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$, (σ_1 and σ_2 unknown, large samples)	$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\left(\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}\right)^{1/2}}$	$(\overline{x}-\overline{y})\pm Z_{\alpha/2}\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}$
6. $\mu_1-\mu_2$ in $N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$ (σ unknown, $n_1+n_2-2>30$)	$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S''\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$	$(\overline{x}-\overline{y})\pm Z_{\alpha/2}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}\; s''$
7. $\mu_1-\mu_2$ in $N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$ (σ unknown, $n_1+n_2-2<30$)	,,	$(\overline{x}-\overline{y})\pm t_{\alpha/2,\, n_1+n_2-2}\sqrt{\dfrac{1}{n_1}\;\dfrac{1}{n_2}}\; s''$
8. σ^2 in $N(\mu, \sigma)$	nS^2/σ^2	$\left(\dfrac{ns^2}{\chi^2_{\alpha/2,\, n-1}},\ \dfrac{ns^2}{\chi^2_{1-\alpha/2,\, n-1}}\right)$
9. σ in $N(\mu, \sigma)$	,,	$\left(\sqrt{\dfrac{ns^2}{\chi^2_{\alpha/2,\, n-1}}},\ \sqrt{\dfrac{ns^2}{\chi^2_{1-\alpha/2,\, n-1}}}\right)$
10. p in a binomial situation with parameter p. (large N, $Np>5$, $N(1-p>5)$	$\dfrac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{N}}}$	$\hat{p}\pm Z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{N}}$
11. p_1-p_2 in two independent binomial situations with parameters p_1 and p_2 (large number of trials).	$\dfrac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\dfrac{\hat{p}(1-\hat{p}_1)}{N_1}+\dfrac{\hat{p}_2(1-\hat{p}_2)}{N_2}}}$	$(\hat{p}_1-\hat{p}_2)\pm Z_{\alpha/2}\times\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{N_1}}\;\sqrt{\dfrac{\hat{p}_2(1-\hat{p}_2)}{N_2}}$

in a population the parameter space, is a subset in a three dimensional space etc.) and the probability that $(\mu, \sigma) \epsilon$ C is $1-\alpha$.

$$P\{(\mu, \sigma) \epsilon C\} = 1-\alpha \tag{8 60}$$

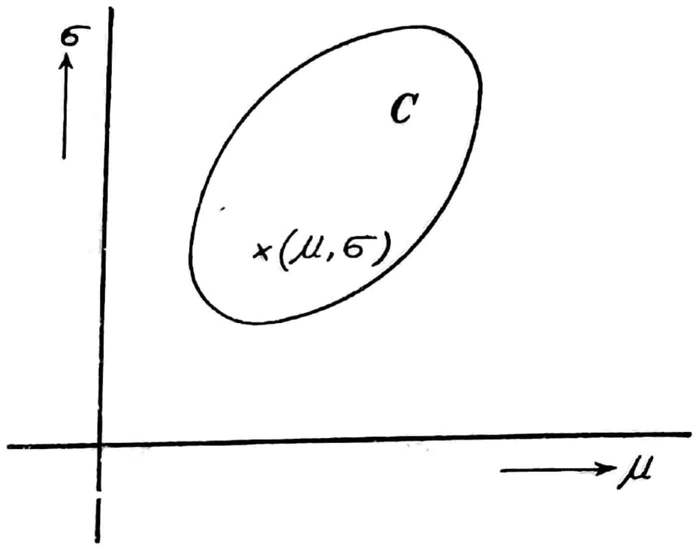


Fig. 8.6

Based on independent samples of size n if $100(1-\alpha)\%$ confidence regions are constructed, in the long run, $100(1-\alpha)\%$ of these regions will cover the true parameter point (μ, σ) These ideas may be extended to the case of any number of parameters and for any parent population. For a more thorough discussion of these topics and other related topics like tolerance limits, fiducial intervals, and Beyesian intervals, the reader is advised to see the Advanced Theory of Statistics, Vol. 2 by M.G. Kendall and A. Stuart. More references are given at the end of this chapter

8.9. CONTROL CHARTS

In industrial production process it is often necessary to check the quality of the product in order to keep the process 'under control'. If pipes of a fixed inner diameter are produced, they may not be good if the diameter goes below a limit or above a certain limit. Examination of each and every item may not be possible if large number of items are produced in short intervals of time. In order to keep the quality of a product within some specified quality limits, industrial engineers use a control chart. Even though control limits, are different from confidence limits it may be easier for the reader to pick up the ideas now. So a note on control charts is given in this section.

8.91. Control Charts for Means. This is a control chart based on the sample means. Suppose that a machine produces a metal rod of length μ_0. If the length is distributed as a $N(\mu_0, \sigma)$ and if we take a random sample of size n then

$$P\{\mu_0 - 3\sigma/\sqrt{n} \leqslant \bar{x} \leqslant \mu_0 + 3\sigma/\sqrt{n}\} = 0.997$$

where $\bar{x}$ denotes the sample mean. If random samples of size n are taken at regular intervals and if an $\bar{x}$ falls outside the interval $(\mu_0 - 3\sigma/\sqrt{n}, \mu_0 + 3\sigma/\sqrt{n})$ then the process may be called 'not under control'. Here $\mu_0 + 3\sigma/\sqrt{n}$ and $\mu_0 - 3\sigma/\sqrt{n}$ are called the upper and lower control limits. A chart is given in Fig. 8.7.

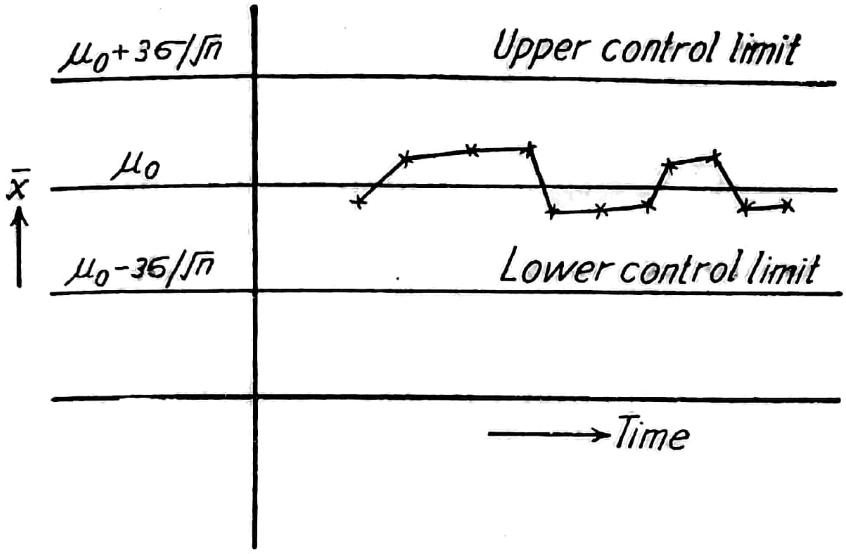


Fig. 8.7.

The sample means of samples of fixed size, taken at regular intervals of time are plotted in the diagram. If the points fall with in the control limits the process is 'under control'. If a point falls outside any of the control limits the fault may be corrected by checking the process. Depending upon the nature of the production processes various control limits can be set up and the production process can be checked with the help of a control chart. The limits in the chart of Fig. 8.7 may be called the 3σ limits, since they are based on a deviation, equal to three times the standard error of $\bar{x}$.

8.92. Control charts for proportions. If $\hat{p}$ is the sample proportion of a sample of size n and if p is the true proportion, in a binomial probability situation then $E(\hat{p}) = p$ and $\text{Var}(\hat{p}) = p(1-p)/n$. So a control chart for proportions may be set up as shown in Fig. 8.8.

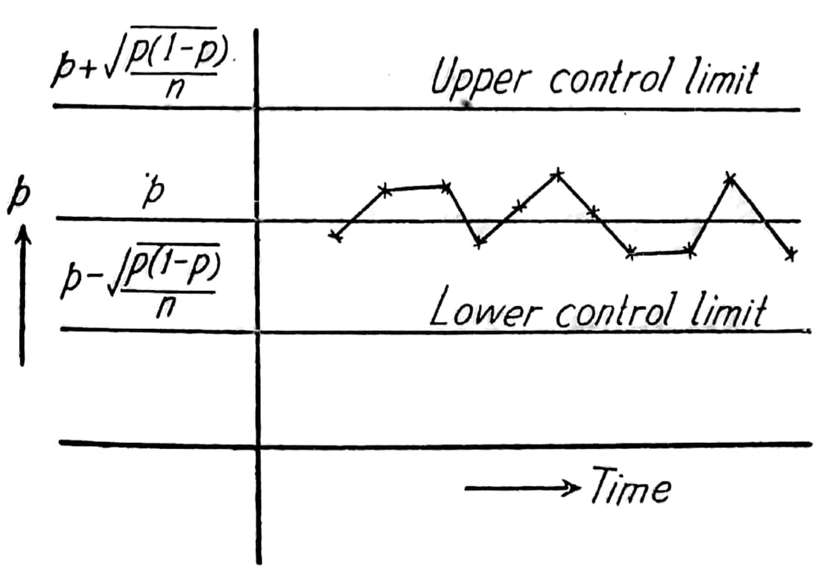


Fig. 8.8

This chart is made by plotting the sample proportions $\hat{p}$ at regular time intervals. If a machine is known to produce $p\%$ defectives, then this percentage may be kept 'under control' by taking a random sample of size n at regular time intervals and plotting the proportion of defectives $\hat{p}$. If $\hat{p}$ falls with in the control limits set up, the production is 'under control'. Here also the control limits may be set up according to the special nature of the production process. The upper and lower control limits in Fig. 8.8 are called the 3σ limits because the limits are set up by considering $p \pm 3\sqrt{\text{Var}(\hat{p})}$. In many processes the lower control limit may not be of any interest to the manufacturer. In that case only the upper limit is considered. Depending upon the nature of the production process one σ, 2σ, or 3σ limits (either both upper and lower limits or only one limit alone) may be set up for checking the quality of goods produced under a particular production process. The same ideas can be used for setting up quality limits for differences of means, for differences of proportions or for variances etc. if desired. If the population is Normal and if a 3σ chart is made then we know that the probability of a sample mean falling outside the control limits is only approximately 0.01, since $P\{\mu - 3\sigma/\sqrt{n} < \bar{x} < \mu + 3\sigma/\sqrt{n}\} = 0.99$ approximately. Even if the population is not normal we know by Chebyshev's inequality that the probability of an $\bar{x}$ falling outside the 3σ control limits is less than 1/9 whatever may be the population. If μ and σ are unknown, they can be replaced by appropriate estimates which can be obtained from the producer's experience or by examining a few samples.

Exercises

8.21. Taking $\mu_0 = 5$ and $\sigma = 0.15$ construct a 2σ and a 3σ control charts for the means. Plot the following data and check whether the process is 'out of control' at any time. The sample means of the samples of size 9 taken at half an hour interval are given below. 4.87, 4.9,5, 5.1, 5.11, 5.13,5, 4·9, 4.85, 4.84, 4.87.

8.22. Taking the true proportion as 0.10 construct a 3σ control chart for the proportion of defectives in a production process and plot the following data. The number of defectives in a random sample of size 100 taken at one hour interval are 5,4,5,7,10,11,10,9,7,8,5,4,2,0,3,5,5,7,8,11,12,14. Comment on the process.

BIBLIOGRAPHY

1. H. **Cramer**. Mathematical Methods of Statistics, Princeton University Press, 1961.

2. D A.S. **Fraser**. Non-parametric Methods in Statistics, New York, John Wiley and Sons Inc. 1957.

3. M.G **Kendall** and A **Stuart**. The Advanced Theory of Statistics, Vol. 2, Charles Griffin and Co. Ltd., London, 1961.

4. E. L **Lehman**. Testing Statistical Hypothesis, New York, John Wiley and Sons Inc. 1959.

5. A. M. **Mood** and F.A. **Graybill**. Introduction to the Theory of Statistics, McGraw-Hill Book Co. Ltd., New York, 1963.

6. C.R. **Rao**. Advanced Statistical Methods in Biometric, Research, New York, John Wiley and Sons Inc, 1952.

CHAPTER 9

POINT ESTIMATION

9.0. Introduction. In the last chapter we considered the problem of giving an interval estimate for a parameter of a probability distribution. In this chapter we will consider point estimates or scalar (single) quantities as estimates of the parameters. In day to day life we face many situations where we would like to get an estimate of a certain unknown quantity. If we want the average weight at a particular time of all teenagers in a particular country, we can find this out if we observe the weight of all the individuals in this population of teenagers. Another method is to select a representative sample and take the sample mean weight as an estimate for the average weight (population mean of the population of the weight measurements). If the weight measurement is distributed as a $N(\mu, 1)$, the problem reduces to the estimation of μ in a $N(\mu, 1)$ from an observed random sample from a $N(\mu, 1)$. If a manufacturer is interested in the average lifetime of electric bulbs manufactured by a particular process, he cannot test each and every bulb, then there won't be any bulb left for sale. He is compelled to obtain an estimate of the average lifetime of electric bulbs by observing a sample (by testing a sample of bulbs). Cost considerations and other numerous factors will compel an experimenter to have an estimate of the parameters. If the lifetime X, of the bulbs is assumed to have an exponential distribution, the problem is to estimate the parameter E(X) in an exponential distribution. If scalar quantities are given as estimates of the parameters, for example, the population mean is estimated by the sample mean or sample median, the population variance estimated by the sample variance etc., such estimates are called point estimates. There are a number of commonly used methods of obtaining point estimates. Some of them are the method of moments, the method of maximum likelihood, the method of least squares, the method of minimum chi-square, Minimax, Invariance and Bayes procedures etc. The method of moments and the method of maximum likelihood will be discussed in the following sections and the other methods will be dealt with later.

9.1. METHOD OF MOMENTS

The motivation for this method is simple and straightforward. We would like to estimate the population moments by the corres-

ponding sample moments. This is one of the oldest methods. The sample moments are equated to the corresponding population moments and the parameters are estimated. That is, by using the equations,

$$m'_r = \mu'_r, \quad r=1, 2, \ldots \tag{9.1}$$

the parameters are estimated, where $m'_r = \Sigma x_i^r / n$ and $\mu'_r = E(X^r)$ and $x_1, \ldots, x_n$ is an observed sample. Of course this method is applicable only if the population moments exist and they are representable in terms of the parameters. For a Cauchy distribution the first moment does not exist and hence this method of estimation cannot be used there.

Ex. 9.1.1. *Estimate the parameters in (1) exponential distribution with the parameter θ, (2) a $N(\mu, \sigma)$, by the method of moments, based on observed random samples of size n.*

Sol. Let $x_1, \ldots, x_n$ be an observed sample. $m'_1 = \Sigma x_i / n = \bar{x}$,

$$m'_2 = \Sigma x_i^2 / n. \tag{9.2}$$

(1) Let $f(x) = \begin{cases} (1/\theta)\, e^{-x/\theta} \\ 0 \text{ elsewhere} \end{cases}$

$E(X) = \theta$ and therefore $m'_1 = \theta$ or $\hat{\theta} = \bar{x}$, where $(\wedge)$ denotes the estimated value. (9.3)

(2) $E(X) = \mu$, $E(X^2) = \mu'_2$, or $\mu'_2 - \mu'^2_1 = \sigma^2$.

The estimating equations are,

$$m'_1 = \mu'_1 \Rightarrow \hat{\mu} = \bar{x}$$

$$m'_2 = \mu'_2 \Rightarrow \hat{\sigma}^2 = m'_2 - m'^2_1 = \Sigma x_i^2 / n - \bar{x}^2 = \Sigma (x_i - \bar{x})^2 / n \tag{9.4}$$

9.2. THE METHOD OF MAXIMUM LIKELIHOOD

Let $x_1, \ldots, x_n$ be an observed random sample of size n from a population $f(x, \theta)$. The joint probability function of $X_1, \ldots, X_n$ (whose observed values are $x_1, \ldots, x_n$) is $f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta) = L(\theta)$ may be considered to be a function of θ. R.A. Fisher who introduced this method, called L, the likelihood function. If the parameters in $f(x, \theta)$ are estimated by maximizing L with respect to the parameters, the method is called the method of maximum likelihood. Opinion is divided on the logical basis of maximizing L with respect

to the parameters which are constants, even though they may be unknown. For a discussion of this topic, the reader may examine the references at the end of this chapter. This method fails if $L(\theta)$ does not have a maximum, and if $L(\theta)$ has a number of local maxima we may take the largest one among them.

Ex. 9.2.1. *Given a random sample $x_1, \ldots, x_n$ from a $N(\mu, \sigma)$ obtain the maximum likelihood estimates of μ and σ.*

Sol. $f(x, \theta) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-(x-\mu)^2/2\sigma^2}$.

$$L = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta) \tag{9.5}$$

$$= (2\pi\sigma^2)^{-n/2} e^{-\Sigma_{i=1}^{n} (x_i-\mu)^2/2\sigma^2} \tag{9.6}$$

$$\log L = -n \log \sigma - (n/2) \log (2\pi) - \Sigma_{i=1}^{n} (x_i - \mu)^2/2\sigma^2 \tag{9.7}$$

$$\frac{\partial}{\partial \mu} \log L = 0 \Rightarrow \Sigma_{i=1}^{n} (x_i - \mu)/\sigma^2 = 0$$

and
$$\frac{\partial}{\partial \sigma} \log L = 0 \Rightarrow -n/\sigma + \Sigma(x_i-\mu)^2/\sigma^3 = 0 \tag{9.8}$$

That is, $\hat{\mu} = \bar{x}$ and $\hat{\sigma} = [\Sigma(x_i - \bar{x})/n]^{1/2}$ (9.9)

It is easily seen that log L is a maximum at $\hat{\mu}$ and $\hat{\sigma}^2$. If $\hat{\mu}$ and $\hat{\sigma}^2$ maximize log L they maximize L also. Hence the maximum likelihood estimates of μ and σ^2 are $\hat{\mu}$ and $\hat{\sigma}^2$ respectively.

Comments. It may be noticed that we need not always take the logarithm of L. The logarithm is taken only for convenience. Obtaining maxima or minima by differentiation may not always be possible. In those cases we will maximize L by using some other methods. If $\hat{\theta}$ is the maximum likelihood estimate of θ and if $\phi(\theta)$ is a non-trivial function of θ then $\phi(\hat{\theta})$ is the maximum likelihood estimate of $\phi(\theta)$. This is easily seen from the following results.

$$\frac{\partial L}{\partial \phi(\theta)} = \frac{\partial L}{\partial \theta} \Big/ \frac{\partial \phi(\theta)}{\partial \theta} = 0 \text{ implies that } \frac{\partial L}{\partial \phi(\theta)} \text{ and } \frac{\partial L}{\partial \theta}$$

vanish together and hence the estimate of $\phi(\theta)$ is $\phi(\hat{\theta})$. So by using this property the maximum likelihood estimate of σ^2 in the above example is $\hat{\sigma}^2 = \Sigma(x_i - \bar{x})^2/n$.

Ex. 9.2.2. *Obtain the maximum likelihood estimate of θ in the following distribution.*

$$f(x, \theta) = \begin{cases} 1/\theta \text{ for } 0 < x < \theta \\ 0 \text{ elsewhere.} \end{cases}$$

Sol. Let $x_1, \ldots, x_n$ be an observed random sample from $f(x, \theta)$,

$$L = 1/\theta^n \text{ and } 0 < x_1, \ldots, x_n < \theta \tag{9.10}$$

Differentiation techniques are not of much help here. If $\hat{\theta}$ is the required estimate for θ, then $\hat{\theta}$ maximizes L, which implies that $\hat{\theta}$ minimizes θ^n, which implies that $\hat{\theta}$ minimizes $\hat{\theta}$. Based on the observations, the smallest possible value which may be assigned to θ, for which L is a maximum, is the largest of the observations, since $0 < x_1, \ldots, x_n < \theta$. Hence $\hat{\theta} = \max(x_1, \ldots, x_n)$ (the largest of the observations.

Comments. If we accept the statement that L is a function of the parameters, the maximum likelihood procedure will be the simple mathematical problem of maximizing L with respect to the parameters. Even though the examples worked out are for the continuous distributions, we can apply the method to discrete as well as to continuous distributions. Here if we use the method of moments, we get $\hat{\theta}$ to be $2\bar{x}$.

Exercises

9.1. Inspection of a random sample of 20 oranges from a large shipment of oranges showed that 2 are spoiled. Obtain a point estimate of the proportion of spoiled ones in the shipment, by the method of moments and by the method of maximum likelihood.

9.2. An office switch board received 2 and 3 phone calls in two randomly selected 5 minute intervals, respectively. Assuming a Poisson distribution for the number of calls in a ten minute interval, time being measured in 5 minute units, obtain a point estimate for the expected number of calls in a 10 minute interval.

9.3. If $x_1, \ldots, x_n$ is a random sample from a Gamma distribution with the parameters α and β, obtain the maximum likelihood estimates of 2α and 3β.

9.4. The income in excess of \$2,000 of the people in a city, is distributed exponentially. Three people, selected at random from this city have incomes \$3000, \$5000 and \$10,000 respectively. Obtain a point estimate of the expected income of a person in this city, by the method of maximum likelihood.

9.5. After the appointment of a new salesgirl the sales in a shop has increased. Four randomly selected days show an increase of \$100, \$300, \$400 and \$600 respectively. Assuming that the increase in sales has an exponential distribution and the provincial tax is 5% of the sales, obtain a point estimate of the expected increase per day of the provincial tax returns from this shop.

9.6. Obtain the point estimates of the parameters in the distribution $f(x, \theta) = 1/(\beta - \alpha)$ for $\alpha < x < \beta$, $\alpha > 0$ and is zero elsewhere, by the method of maximum likelihood. (Assume that a random sample is given).

9.7. If $f(x) = (\theta + 1)\, x^\theta$ for $0 < x < 1$, $\theta > 0$ and is zero elsewhere. Obtain a point estimate of θ by the method of moments and by the method of maximum likelihood. (Assume that a random sample of size n is given).

PROPERTIES OF ESTIMATORS

It is seen that the different estimation procedures discussed above can lead to different estimates to the same parameter. If we estimate the same parameter using two different samples we may get different estimates. Owing to the arbitrariness of the methods discussed, one may argue that an observed value of any statistic may be taken as an estimate for a parameter. In general, any statistic, if it is used to estimate a parameter, may be called an estimator. In the following sections we will formulate some desirable properties of estimators, so that we will be able to say that one estimate is better than another estimate or to show that one particular estimate is the best of all possible estimates, having some general properties. Since there is no unique set of criteria by which one can select anestimator which is the best among all possible estimators, we will discuss some general criteria which are desirable under some conditions. The basis of preferring one estimator to another depends upon the purpose for which the estimator is used. In a particular case one estimator may be preferable to another estimator satisfying more properties. So in the following sections a number of criteria will be discussed and the desirable ones for a particular situation can be picked up by studying the conditions and the purpose for which the estimator is used. In other words the selection of a particular estimator is left to the experimenter who wants to use an estimator and who knows his experimental conditions.

9.3. UNBIASEDNESS

If $\hat{\theta}$ is an estimator of θ such that $E(\hat{\theta})=\theta$ then $\hat{\theta}$ is called an unbiassed estimator and the value assumed by $\hat{\theta}$ is called an unbiased estimate for θ. (Whenever there is no confusion we will use $\hat{\theta}$ for the estimator as well as for the estimate). We know that $E(\bar{X})=\mu$ and hence the sample mean is an unbiased estimator of the population mean whatever may be the population. If $x_1,\ldots, x_n$ is an observed sample then $\bar{x}=(x_1+\cdots+x_n)/n$ may be considered to be a value assumed by $\bar{X}$ and hence $\bar{x}$ is called an unbiased estimate of μ. But we know that $E(S^2)=E\ \Sigma\ (X_i-\bar{X})^2/n\neq\sigma^2$. Hence the sample variance is not an unbiased estimator for the population variance. But $E\ \Sigma(X_i-\bar{X})^2/(n-1)=\sigma^2$ and therefore $\Sigma\ (X_i-\bar{X})^2/(n-1)$ is an unbiased estimator for σ^2, whatever may be the population. Here the bias in an estimator S^2 is easily removed by multiplying the estimator by a constant. This procedure is not applicable if the expected value of the estimator is a complicated function of the parameter. In this case some general methods for removing the bias are available (see reference [3] at the end of this chapter). Unbiasedness is a desirable property. If we go on taking random samples of the same size, we would like our estimator to assume the parameter, on the average, in the long run. However sometimes from other considerations, unbiasedness may not be desirable. If there

exists, a statistic T such that $E(T)=g(\theta)$ then $g(\theta)$ is said to be an estimable parametric function.

Ex. 9.3.1. *Show that in a binomial probability situation the estimate of the probability of a success p, by the method of maximum likelihood, is unbiased for p.*

Sol. If $X_1, \ldots, X_n$ denote the independent stochastic variables taking the values 1 with probability p and 0 with probability $(1-p)$ and other values with zero probabilities, then the number of successes X in N independent trials is,

$$X=X_1+\ldots+X_N$$

The maximum likelihood estimate of p is easily seen to be,

$$\hat{p}=x/N=\text{observed proportion of successes.} \tag{9.11}$$

Therefore, the estimator is X/N and $E(X/N)$

$$=(1/N)E(X)=Np/N=p. \tag{9.12}$$

Hence X/N is an unbiased estimator of p or $\hat{p}=x/N$ is an unbiased estimate of p.

9.4. CONSISTENCY

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimates of θ, we will need further criteria in order to call one better than the other. Another desirable property of an estimator is called consistency or stochastic convergence. If the probability of $\hat{\theta}$ tendingt o θ approaches one when n tends to infinity, $\hat{\theta}$ is called a consistent estimator of θ.

That is,

$$P\{\hat{\theta}\to\theta\}\to 1 \text{ as } n\to\infty \tag{9.13}$$

or corresponding to any given $\epsilon > 0$, however small it may be, there exists a $\delta > 0$ such that,

$$P\{\,|\,\hat{\theta}-\theta\,|\,>\epsilon\}<\delta \tag{9.14}$$

for $n \geqslant$ some specified value n_0. This is the same as saying that θ converges to θ in probability.

For example if we want to estimate the population mean of a finite population of size N and if we take a random sample of size n, the sample mean coincides with the population mean when $n=N$. When the sample size n approaches N we would expect the sample mean $\bar{x}$ to approach the population mean μ. In general, in an infinite population when $n\to\infty$ we would like to have the probability of our estimate coinciding with the parameter approximately equal to one. It may be noticed that 'consistency' is a large sample concept or here we are dealing with the property of the estimator when n is very large.

If t is a stochastic variable, by Chebyshev's inequality (see section 3.5).

$$P\{\,|\,t-E\,(t)\,|\,>k\}<\text{Var}\,(t)/k^2. \tag{9.15}$$

Therefore we can give the following theorem.

Theorem 9.1. If $\hat{\theta}$ is an unbiased estimator of θ and if $\text{Var}(\hat{\theta})\to 0$ and $n\to\infty$, then $\hat{\theta}$ is consistant for θ.

Proof. By using the inequality (9.15)

$P\{|\hat{\theta}-\theta|>k\}<\text{Var}(\hat{\theta})/k^2$ where k is any arbitrary positive constant. But $\text{Var}(\hat{\theta})\to 0$ as $n\to\infty$. Therefore when $n\to\infty$, $P\{|\hat{\theta}-\theta|>k\}\to 0$ or $\hat{\theta}$ converges to θ in probability. (9.16)

If $\hat{\theta}$ is a consistent estimator of θ, $n\hat{\theta}/(n+1)$ is also consistent for θ. A number of consistent estimators may be constructed. It may be noticed that a consistent estimator need not be unbiased. For example if $\hat{\theta}$ is a consistent and an unbiased estimator of θ then $n\hat{\theta}/(n+2)$ is not unbiased but is a consistent estimator of θ.

Ex. 9.4.1. *Show that the sample mean is an unbiased and consistent estimator of the population mean of any population having a finite variance.*

Sol.
$$\begin{aligned} E(\bar{X}) &= (1/n)\, E(X_1+\ldots+X_n) \\ &= (1/n)[E(X_1)+\ldots+EX_n)] \\ &= (1/n)\,(\mu+\ldots+\mu)=\mu \text{ (implies unbiasedness)} \end{aligned} \tag{9.17}$$

$\text{Var}(\bar{X})=\sigma^2/n$ where σ^2 is the population variance (see Ex. 5.6.1.) (9.18)

$\text{Var}(\bar{X})\to 0$ as $n\to\infty$. Therefore by theorem 9.1, $\bar{X}$ is a consistent estimator of μ.

9.5. RELATIVE EFFICIENCY

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased and consistent estimators of θ, we need more criteria in order to select the better one. If $\text{Var}(\hat{\theta}_1)>\text{Var}(\hat{\theta}_2)$. we would prefer the one which has smaller dispersion and we will choose $\hat{\theta}_2$ to $\hat{\theta}_1$. Since $\text{Var}(\hat{\theta}_1)$ and $\text{Var}(\hat{\theta}_2)$ are measures of dispersion of $\hat{\theta}_1$ and $\hat{\theta}_2$ from θ respectively, we will base our next criterion on the dispersions of the estimators for the parameter. The relative efficiency of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is defined as

$$e=E(\hat{\theta}_2-\theta)^2/E(\hat{\theta}_1-\theta)^2 \tag{9.19}$$

where $\hat{\theta}_1$ and $\hat{\theta}_2$ are two estimators of θ and E denotes 'mathematical expectation'. If $E(\hat{\theta}_1)=\theta=E(\hat{\theta}_2)$ then

$$e=\text{Var}(\hat{\theta}_2)/\text{Var}(\hat{\theta}_1) \tag{9.20}$$

If $e>1$ then $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$. If $\hat{\theta}$ is an estimator of θ such that $E(\hat{\theta})=\theta$ and $\hat{\theta}$ has a variance smaller than that of any other unbiased estimator of θ then $\hat{\theta}$ is called a 'minimum variance unbiased estimator'. If $e>1$ when the sample size tends to infinity then $\hat{\theta}_1$ can be called 'asymptotically more efficient' than $\hat{\theta}_2$. Again asymptotic efficiency is a large sample concept. For example the sample mean and the sample median are unbiased

estimators of the population mean in a normal population $N(\mu, \sigma)$. But Var $(\overline{X}) = \sigma^2/n$ and Var $(m) = \pi\, \sigma^2/2n$ (see section 6.4), where m is the sample median. Since Var $(\overline{X}) <$ Var (m), $\overline{X}$ is more efficient than m. We will state the following theorem without proof.

Theorem 9.2. If $\hat{\theta}$ is an unbiased estimator of θ and if $\frac{\partial}{\partial \theta} \log L = k(\hat{\theta} - \theta)$, where k is independent of the sample observations, but may be a function of θ, then $\hat{\theta}$ is the minimum variance unbiased estimator of θ, where L is the likelihood function.

Ex. 9.5.1. *Show that the sample mean is the minimum variance unbiased estimator of μ in $N(\mu, \sigma)$.*

Sol. Let $x_1, \ldots, x_n$ be an observed sample from a $N(\mu, \sigma)$

$$L = (2\pi\sigma^2)^{-n/2}\, e^{-\Sigma_{i=1}^{n} (x_i - \mu)^2/2\sigma^2}.$$

$$\log L = -n \log \sigma - (n/2) \log (2\pi) - \Sigma_{i=1}^{n} (x_i - \mu)^2/2\sigma^2$$

$$\frac{\partial}{\partial \sigma} \log L = \Sigma_{i=1}^{n} (x_i - \mu)/\sigma^2 = n(\bar{x} - \mu)/\sigma^2 \tag{9.21}$$

That is, $\frac{\partial}{\partial \mu} \log L = k(\bar{x} - \mu)$ where $k = n/\sigma^2$ which is independent of $x_1, \ldots, x_n$. Also we know that $E(\overline{X}) = \mu$. Hence $\overline{X}$ is the minimum variance unbiased estimator of μ in a $N(\mu, \sigma)$.

Comments. We know that Var $(\overline{X}) = \sigma^2/n$. But $\overline{X}$ is the minimum variance estimator and hence any unbiased estimator of μ has a larger variance. It may be noticed that $1/k = \text{Var}(\overline{X}) = \sigma^2/n$. In general if $\frac{\partial}{\partial \theta} \log L = k\,(\hat{\theta} - \theta)$ where $E(\hat{\theta}) = \theta$ then $1/k = \text{Var}\,(\hat{\theta})$.

9.6. SUFFICIENCY

Another desirable property of an estimator is sufficiency. A statistic $\hat{\theta}$ is called sufficient for θ if the conditional joint distribution of sample values, given $\hat{\theta}$, is independent of θ. That is, $f(x_1, \ldots, x_n/\hat{\theta})$ is independent of θ, where $X_1, \ldots, X_n$ denote the random sample under consideration and if the conditional distribution of $X_1, \ldots, X_n$ given $\hat{\theta}$ is independent of θ, we may say that $\hat{\theta}$ contains all relevant information in the sample about the parameter θ. We may also define a sufficient statistic as that statistic which contains all relevant information about the parameter in the sample. But the likelihood function $L = f(x_1, \ldots, x_n) = f_1(x_1, \ldots, x_n/(\hat{\theta}) \cdot f_2(\hat{\theta})$. If the likelihood function L can be factorized into two functions where one is a function of $\hat{\theta}$ and θ and the other is independent of θ then $\hat{\theta}$ is sufficient for θ. All these three definitions are one and the same. Hence a sufficient statistic provides a reduction in data. Instead of considering all

the sample values we need consider only a sufficient statistic as far as the 'information' about the parameter is concerned.

Ex. 9.6.1. *In a $N(\mu, 1)$ show that the sample mean is a sufficient estimator of μ.*

Sol. $L=(2\pi)^{-n/2}\, e^{-\Sigma_{i=1}^{n} (x_i-\mu)^2/2}$

$$=(2\pi)^{-n/2}\, e^{-\Sigma_{i=1}^{n} (x_i-\bar{x})^2/2}\, e^{-n(\bar{x}-\mu)^2/2} \tag{9.22}$$

But $(2\pi)^{-n/2}\, e^{-\Sigma_{i=1}^{n} (x_i-\bar{x})^2/2}$ is independent of μ and $e^{-n(\bar{x}-\mu)^2/2}$ is a function of $\bar{x}$ and μ. Hence $\bar{X}$ is sufficient for μ since L is factorized into two functions where one is a function of the estimate and the parameter and the other is independent of the parameter.

This can also be demonstrated by showing that the conditional distribution of $X_1,\ldots,X_n$ given $\bar{X}$ is independent of μ. The distribution of the sample mean of a random sample of size n from a $N(\mu, 1)$ is

$$f_2(\bar{x})=(n/2\pi)^{1/2}\, e^{-n(\bar{x}-\mu)^2/2} \tag{9.23}$$

Therefore $f_1(x_1,\ldots x_n \mid \bar{x})=L/f_2(x)=(2\pi n)_{-(n-1)/2}\, e^{-\Sigma(\bar{x}_i-x)^2/2}$ which is independent of μ.

Comments. From this example it may be noticed that if $\bar{X}$ is sufficient for μ, $2\bar{X}$ is also sufficient for μ, since L can be factorized into f_1 and f_2 where f_2 is a function of $2\bar{x}$ and μ and f_1 is independent of μ. In general if $\hat{\theta}$ is sufficient for θ then any one-to-one function of $\hat{\theta}$ is also sufficient for θ. $L=f_1(\hat{\theta}, \theta)\, f_2$ where f_2 is independent of θ. If L is maximized with respect to $\hat{\theta}$ it is equivalent to maximizing $f_1(\theta, \theta)$ since f_2 is independent of θ. Hence the maximum likelihood estimate of θ will be a function of θ, (that is, a function of a sufficient estimate for θ). So we may state the following theorem.

Theorem 9.3. If a single sufficient statistic $\hat{\theta}$ for θ exists, then the maximum likelihood estimate of θ will be a function of $\hat{\theta}$.

9.6.1. Joint Sufficiency. If there are k parameters $\theta_1,\ldots,\theta_k$ in a distribution and if there exist/statistics $t_1,\ldots, t_s$ such that,

$$L=f_1(t_1,\ldots, t_s, \theta_1,\ldots, \theta_k)\, f_2 \tag{9.25}$$

where L is the likelihood function, f_1 is a function of the statistics and the parameters and f_2 is independent of the parameters, then $t_1,\ldots, t_s$ are said to be jointly sufficient for $\theta_1,\ldots,\theta_k$. Since s need not be equal to k, joint sufficiency need not imply that t_1 is

sufficient for θ_1, t_2 is sufficient for θ_2 etc. Even if $k=s$ joint sufficiency need not imply individual sufficiency. If there exists a minimum number r of statistics $t_1, \ldots, t_r$ such that they are jointly sufficient for $\theta_1, \ldots, \theta_k$ then $t_1, \ldots, t_r$ is called a minimal set of sufficient statistics for $\theta_1, \ldots, \theta_k$. Since there can exist a number of sets of sufficient statistics for the parameters, 'smallest' is used in the sense that this minimal set is a function of all other sets of sufficient statistics. In other words, further reduction of data is not possible without losing sufficiency. Incidently it may be noted that the sample itself forms a set of sufficient statistics for the parameters, in any case.

Ex. 9.61.1. *Show that the sample mean and the sample variance are jointly sufficient for μ and σ^2 in a $N(\mu, \sigma)$.*

Sol. $L=(2\pi\sigma^2)^{-n/2}\, e^{-\Sigma(x_i-\mu)^2/2\sigma^2}$

$$=(2\pi\sigma^2)^{-n/2}\, e^{-\Sigma(x_i-\bar{x})^2/2\sigma^2-n(\bar{x}-\mu)^2/2\sigma^2} \tag{9.26}$$

$$=f_1(s^2, \bar{x}, \sigma^2, \mu)\, f_2 \tag{9.27}$$

where $f_2=(2\pi)^{-n/2}$ and f_1 is a function of μ, σ^2, $\bar{x}$ and s^2.

Hence the result.

Comments. f_2 need not be taken as $[1/(2\pi)^{n/2}]$. f_2 may be arbitrarily fixed according to the definition. The only restriction is that f_1 should be a function of s^2, $\bar{x}$, μ and σ^2 and f_2 should not contain μ and σ^2.

As sufficient statistics and maximum likelihood estimators are very useful in statistical analysis, we will state a few theorems without proofs, for the information of the reader. In the following discussion we consider only a single parameter case and the parameter will be denoted by θ, where $\theta \in \Omega$ (parameter space).

Theorem 9.4. If $\log f(x, \theta)$ is differentiable with respect to θ in an interval containing the true θ where $f(x, \theta)$ is the probability function under consideration, then the maximum likelihood estimator is consistent for θ.

Theorem 9.5. Under some general regularity conditions which will be stated below, the maximum likelihood estimator is asymptotically (that is, when the sample size $n \to \infty$) more efficient, sufficient and normally distributed.

Regularity conditions. These are some general and reasonable conditions on the probability function $f(x, \theta)$ under consideration. Let, $\frac{d}{d\theta} f(x, \theta)=f'(x, \theta)$, $\frac{d^2}{d\theta^2} f(x, \theta)=f''(x, \theta)$ and let E denote 'mathematical expectation'.

Condition 1. The derivatives $f'(x, \theta)$, $f''(x, \theta)$ and $f'''(x, \theta)$ exist for almost all values of x in an interval I of θ containing the true value of θ.

Condition 2. At the true value of θ, let, $E[f'(X, \theta)/f(X,\theta)] = 0 = E[f''(X, \theta)/f(X, \theta)]$ and $E[(f'(X, \theta))^2/f(X, \theta)] > 0$.

Condition 3. For every θ in I, $\left| \frac{d^3}{d\theta^3} \log f(x, \theta) \right| < M(x)$ and $E\, M(X) < k$ where k is independent of θ.

9.7. COMPLETENESS

This is basically a property of families of probability functions. This was introduced in section 3.6. But here we will consider the ideas of complete statistics or complete estimators and complete sufficient estimators. In section 9.3 we discussed the concept of 'unbiasedness of an estimator. If T_1 and T_2 are two unbiased estimators for a function $h(\theta)$ of θ then $ET_1 = h(\theta) = ET_2$ and $E(T_1 - T_2) = 0$. For any two statistics T_1 and T_2, if $E(T_1 - T_2) = 0$ implies $T_1 - T_2 = 0$ almost everywhere then $T_1 - T_2 = 0$ is the only estimator for zero. This induces a uniqueness for the unbiased estimator for $h(\theta)$. Incidently if a statistic T exists such that $E(T) = h(\theta)$ then $h(\theta)$ is said to be an estimatable parametric function. If for any statistic T with probability function $f(t, \theta)$ where $\theta \in \Omega$ (parameter space), $E\, g(T) = 0$ for all $\theta \in \Omega$, implies that $g(t) = 0$ almost everywhere then T is called a complete statistic, where $g(T)$ is a real function of T and E denotes 'mathematical expectation'. If T is sufficient also then T is a complete sufficient statistic.

Ex. 9.7.1. *Show that the sample mean of a random sample from a $N(\mu, 1)$ is a complete sufficient estimator for μ.*

Sol. Sufficiency of the sample mean $\bar{X}$ was seen in Ex. 9.6.1. $\bar{X} : N(\mu, 1/\sqrt{n})$

$$f(\bar{x}) = (n/2\pi)^{1/2}\, e^{-n(\bar{x}-\mu)^2/2}$$

$$-\infty < \bar{x} < \infty, \text{ and } \mu \in \Omega$$

where $\Omega = \{y \mid -\infty < y < \infty\}$.

Let $g(\bar{X})$ be a function of $\bar{X}$ and let $E\, g(\bar{X}) = 0$.

$$(n/2\pi)^{1/2} \int_{-\infty}^{\infty} g(\bar{x})\, e^{-n(\bar{x}-\mu)^2/2}\, d\bar{x} = 0 \tag{9.29}$$

$$\Rightarrow \quad \int_{-\infty}^{\infty} g(\bar{x})\, e^{-n(\bar{x}-\mu)^2/2}\, d\bar{x} = 0 \tag{9.29}$$

$$\Rightarrow \quad \int_{-\infty}^{\infty} g(\bar{x})\, e^{-n\bar{x}^2/2} \quad e^{n\mu\bar{x}} = 0 \text{ (since } e^{n\mu^2/2} \neq 0) \tag{9.30}$$

$\Rightarrow \quad g(\bar{x})\, e^{-n\bar{x}^2/2} = 0$ (This is obtained by taking a Laplace transform. See problems 4.51 and 4.52) (9.31)

$$\Rightarrow \quad g(\bar{x})=0 \text{ (since } e^{n-\bar{x}^2/2}\neq 0) \tag{9.32}$$

Comments. A reader who is not familiar with Laplace and Fourier transforms may omit this section. Since many proofs of completeness involve complicated transformations this section will not be elaborated.

9.8. INVARIANCE

This is a desirable property for the estimators. A rigorous definition and complete explanation of 'Invariance' involves Mathematics beyond the pre-requisite of this book. So only the basic ideas are introduced through some examples in this section. Consider the following problem of estimation. A population mean μ is estimated by using an observed random sample $x_1,\ldots, x_n$. A number of estimates can be given for μ. Let $T(x_1,\ldots, x_n)$ be an estimate for μ. Suppose that the observations are measurements in feet and we would like to have the estimate in inches. The new observations are $12x_1,\ldots, 12x_n$ in inches and the estimate is $T(12x_1,\ldots, 12x_n)$. A very desirable property of the estimate in this problem is

$$T(12x_1, \ldots, 12x_n)=12T(x_1,\ldots, x_n),$$

since μ feet$=12\mu$ inches. Consider two estimators $T_1(X_1,\ldots, X_n)=\Sigma X_i/n$ and $T_2(X_1,\ldots, X_n)=\Sigma X_i^2/n$ for μ T_1 satisfies the invariance requirement since $T(aX_1\ldots, aX_n)=a\,T(X_1,\ldots, X_n)$ for every scalar a. But $T_2(aX_1, \ldots, aX_n)\neq aT_2(X_1,\ldots, X_n)$ and hence T_2 does not satisfy the invariance requirement with respect to a scale transformation. If an estimator $T(X_1,\ldots, X_n)$ for μ is such that $T(X_1+c,\ldots, X_n+c)=c+T(X_1,\ldots, X_n)$ for all real c then T, as an estimator for μ, satisfies the invariance requirement with respect to a translation or a change in the location. It can be seen that the sample mean, as an estimator for the population mean, satisfies the invariance requirements with respect to a scale as well as location transformation. In general we can define the invariance property of an estimator for a parametric function, with respect to some general transformations satisfying some conditions. This also leads to a method of estimation called the 'invariance method' by which one can get estimators satisfying the invariance properties. This will not be discussed here. In the invariance problems, even though the estimates from the original observations and from the transformed observations may be different, the structure of the estimation procedure remains the same, in the sense that the family of distributions remains the same. This is why the procedure is called the 'invariance method'.

Exercises

9.8. Obtain an unbiased estimator of p^2+2 where p is the parameter of the binomial distribution,

$$f(x, \theta)=\begin{cases}\binom{N}{x} p^x (1-p)^{N-x} & \text{for } x=0, 1, \ldots, N\ ;\ 0<p<1 \\ & \text{N is known} \\ 0 \text{ elsewhere.} \end{cases}$$

[**Hint :** Obtain the second factorial moment].

9.9. If $x_1, \ldots, x_n$ is a random sample from a Poisson distribution with parameter λ, show that $\bar{X}$ is an unbiased and minimum variance estimator of λ.

9.10. If $\bar{X}$ is the sample mean of a random sample from a $N(\mu, 1)$ show that $3\bar{X}+2$ is a sufficient statistic for μ.

9.11. If $\bar{X}$ and S^2 are the mean and variance of a random sample of size n from a $N(\mu, \sigma)$ show that $2\bar{X}$ and $3S^2$ are jointly sufficient for μ and σ^2.

9.12. If $\hat{\theta}$ is a consistent estimator of θ and if N is the sample size, show that $(N-a)\,\hat{\theta}/(N-b)$ is also consistent for θ, where a and b are constants.

9.13. In a Cauchy distribution,

$$f(x, \theta)=1/\pi \{1+(x-\theta)^2\}, \quad -\infty<x<\infty$$

show that the sample mean is not a consistent estimator of θ.

9.14. Obtain the minimum variance unbiased estimator of p, if it exists, in a binomial distribution,

$$f(x, \theta)=\begin{cases}\binom{N}{x} p^x (1-p)^{N-x} & \text{for } x=0, 1, \ldots, N\ ;\ 0<p<1. \\ & \text{N is known.} \\ 0 \text{ elsewhere.} \end{cases}$$

9.15. If X_1, X_2, X_3, X_4 is a random sample of size 4 from a Poisson distribution with parameter λ, show that, $\hat{\theta}_1=(X_1+X_2+X_3+X_4)/4$ and $\hat{\theta}_2=(2X_1+3X_3)/5$, are both unbiased. Which one is relatively more efficient ? Are they sufficient for λ ?

9.16. Show that the sample mean is an unbiased and consistent estimator of $\theta+\frac{1}{2}$ for the following distribution.

$$f(x)=\begin{cases} 1 \text{ for } \theta<x<\theta+1 \\ 0 \text{ elsewhere.} \end{cases}$$

9.17. If L is the likelihood function when a random sample of size n is taken from a population $f(x, \theta)$ having a parameter θ, then

$$I=E\left(\frac{\partial}{\partial\theta} \log L\right)^2$$

is sometimes called the amount of information about θ, in the sample. Show that

$$I=E\left(\frac{\partial}{\theta} \log L\right)^2=-E\left(\frac{\partial^2}{\partial\theta^2} \log L\right)$$

$$=n\, E\left[\frac{\partial}{\partial\theta} \log f(x)\right]^2$$

(Assume the regularity conditions mentioned in section 9.6).

9.18. Under the assumptions in problem 9.17 and using the result that

$$[\mathrm{Cov}\,(X, Y)]^2 \leqslant \mathrm{Var}\,(X).\ \mathrm{Var}\,(Y), \text{ show that,}$$

$$\mathrm{Var}\,(\hat{\theta}) \geqslant 1/I$$

where I is defined in problem 9.17 and $\hat{\theta}$ is an unbiased estimator of θ in a population $f(x, \theta)$. This inequality is called the Cramer-Rao inequality. More general forms of the inequality can be given.

9.19. If $E(\hat{\theta})=\theta+b(\theta)$ where $b(\theta)$ is a function of θ and if

$$b'(\theta)=\frac{d}{d\theta}\,b(\theta), \text{ show that}$$

$$\mathrm{Var}\,(\hat{\theta}) \geqslant [1+b'(\theta)]^2/I$$

9.20. If $E(\hat{\theta})=\theta$ and $\dfrac{\partial}{\partial\theta}\log L=k\,(\hat{\theta}-\theta)$ show that $\hat{\theta}$ is the minimum variance unbiased estimator of θ (see theorem 9.2).

9.21. Show that the statistic X_1+X_2 is complete and sufficient for θ, where X_1, X_2 is a random sample of size 2 from a population defined as follows.

$$f(1)=\theta, f(0)=1-\theta \text{ and } f(x)=0 \text{ elsewhere, where } 0<\theta<1.$$

9.22. Show that the sample mean, as an estimator for the population mean μ, under the transformation $x \xrightarrow{g} y=ax+b,\ a>0,\ -\infty<b<\infty$, (that is whenever there is an observation x we consider a new quantity $y=ax+b$ and this transformation is denoted by g), satisfies the invariance requirements.

9.23. Give an estimator T for the parameter θ of an exponential distribution, which is not (1) unbiased, (2) sufficient, (3) complete, (4) satisfying the invariance requirements under a change in the scale.

BIBLIOGRAPHY

1. H. **Cramer,** Mathematical Methods of Statistics. Princeton University Press, 1961.

2. D.A.S. **Fraser**. Non-Parametric Methods in Statistics. John Wiley, New York, 1957.

3. M.G. **Kendall** & A. **Stuart.** The Advanced Theory of Statistics. Vol. 2, Charles Griffin and Co. London, 1961.

4. A.M. **Mood** & F.A. **Graybill,** Introduction to the Theory of Statistics. McGraw-Hill Book Co, Ltd., New York, 1963.

5. C.R. **Rao,** Advanced Statistical Methods in Biometric Research. John Wiley, New York, 1952.

6. C.R. **Rao,** Linear Statistical Inference and Its Applications. John Wiley, New York, 1965.

CHAPTER 10

TEST OF HYPOTHESES

10.0. Introduction. In the last two chapters we discussed some problems of statistical inference, namely, estimation of parameters and setting up of confidence intervals. In this chapter we shall consider the problem of testing statistical hypotheses. This remarkable aspect of statistical theory had led some people to claim that anything and everything can be proved by statistics. By using statistical methods we do not prove anything but these methods help us to make decisions in situations where there is a lack of certainty. There are many practical situations where we would like to take a decision for further action. There are other problems where we would like to determine whether some specific claims are acceptable or not. Suppose that we want to test the following claims.

1. A particular toothpaste reduces cavities by 46%.

2. A particular drug raises the survival rate from a disease to 95%.

3. The appointment of a new salesgirl increases average sales by \$200 a day.

4. A bird watcher claims that during the spring season on the average, birds lay more eggs in Quebec shan in Ontario.

5. Detergent A is more powerful than detergent B.

6. The fertilizers F_1, F_2, F_3 and F_4 are equally effective as far as the yield of a particular variety of corn is concerned.

7. Detergent A out cleans all other detergents.

These are a few of the many varieties of problems whose solutions demand the help of a statistician. The first problem is testing the hypothesis that a binomial probability is 0.46. The second problem may also be considered to be a binomial probability situation. If the increase in sales in problem 3 is assumed to be normally distributed with mean μ and with a known variance then it is a problem of testing the hypothesis $\mu=\mu_0=200$. Similarly problem 4 may be transformed to a problem of testing the equality of the means in two populations. A statistical hypothesis of this nature is only a restriction on the estimable parameters (parameters for which unbiased estimators exist) of a

probability distribution. Some non-parametric hypotheses (which are not restrictions on estimable parameters) will be considered in the last chapter. In problem 1 when the hypothesis that the true proportion is 0.46 is tested it is tested against an alternative that either the true proportion is less than 0.46 or greater than 0.46 or not equal to 0.46. So a statistical hypothesis and the alternative hypothesis may be formulated as follows.

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta = \theta_1 \tag{10.1}$$

where H_0 (we may call it the null hypothesis) is the hypothesis to be tested against the alternative H_1.

10.1. SIMPLE AND COMPOSITE HYPOTHESES

If a null hypotheses completely specifies a distribution (that is, the functional form as well as the parameters) then it is called a simple hypothesis ; otherwise it is called composite. For example if we want to test the hypothesis $H_0 : \mu = 5$ in a normal population $N(\mu, 1)$ then H_0 is a simple hypothesis. If the alternative is $H_1 : \mu = 10$ then the alternative is also simple. If the alternative is $\mu < 5$ then there are a number of possible values for μ in H_1 and hence the alternative is composite. If $H_0 : \mu > 5$ then H_0 is composite. If H_0 is $\mu = 5$ in a $N(\mu, \sigma)$ where σ is unknown, again H_0 is composite since H_0 does not completely determine the population.

$H_0 : \theta = \theta_0$ The null hypothesis is simple and the alternative
$H_1 : \theta = \theta_1$ is also simple. (10.2)

$H_0 : \theta = \theta_0$ H_0 simple and H_1 composite and one sided.
$H_1 : \theta < \theta_0$ (10.3)

$H_0 : \theta = \theta_0$ H_0 simple and H_1 composite and one sided
$H_1 : \theta > \theta_0$ (10.4)

$H_0 : \theta = \theta_0$ H_0 simple and H_1 composite and two sided
$H_1 : \theta \neq \theta_0$ (10.5)

where θ_0 and θ_1 are specific values of θ and θ denotes the parameter in a given population.

10.2. TYPE I AND TYPE II ERRORS

When a hypothesis H_0 is tested against an alternative H_1 usually there can arise one of the two types of errors, namely to reject H_0 when H_0 is true and to accept H_0 when H_1 is true. These are called the type I and type II errors respectively. They are illustrated in the following table. Here we will assume, for the time being, that rejection of H_0 is equivalent to acceptance of H_1 and *vice versa*. For example if $H_0 : \mu = 20$ in a $N(\mu, 1)$ is rejected

in favour of the alternative $H_1 : \mu=30$ then automatically H_1 is accepted.

	H_0 is true (H_1 is not true)	H_0 is not true (H_1 is true)
Accept H_0 (reject H_1)	Correct decision	Type II error
Reject H_0 (accept H_1)	Type I error	Correct decision

The acceptance and rejection of H_0 depend on the test criterion that is used in testing H_0. In a particular testing procedure we would like to control the type I as well as the type II error. The probabilities of committing the type I and type II errors are called the sizes of the two errors and are denoted by α and β respectively.

Type I error size α

Type II error size β

In the following examples we shall consider the problem of evaluating type I and type II errors if a criterion for testing the null hypothesis H_0 is given.

Ex. 10.2.1. *The consumption of electricity in a small township is assumed to be exponentially distributed with parameter θ. Determine the sizes of the type I and type II errors if $H_0 : \theta=1000$ k.w. is tested against the alternative $\theta=2000$ and if the test criterion is as follows. Select any day at random. If the consumption on that day is 4000 k.w. or more reject H_0 otherwise accept H_0.*

Sol. The distribution is

$$f(x, \theta)=\begin{cases} \dfrac{1}{\theta}\, e^{-x/\theta} & \text{for } x>0,\ \theta>0 \\ 0 & \text{elsewhere} \end{cases}$$

According to the test criterion, H_0 is rejected if $x \geqslant 4000$

$\therefore$ α = probability of rejecting H_0 when H_0 is true

= ,, ,, when $\theta=1000$

$$=\int_{4000}^{\infty} \frac{1}{1000}\, e^{-x/1000}\, dx = e^{-4000/1000} = e^{-4}$$

β=probability of accepting H_0 when H_1 is true

= ,, ,, ,, when $\theta=2000$.

But H_0 is accepted when $x<4000$ and hence

$$\beta=\int_0^{4000} \frac{1}{2000} e^{-x/2000} dx=1-e^{-4000/2000}=1-e^{-2}$$

The two probabilities are illustrated in Fig. 10.1.

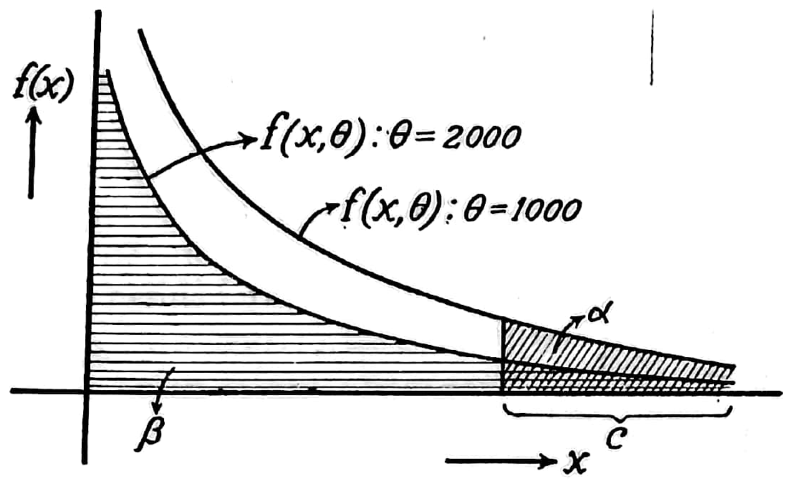


Fig. 10.1.

Ex. 10.2.2. *A coin is thrown 10 times. Suppose that the hypothesis H_0 : $p=1/2$ is rejected in favour of the alternative H_1 : $p=2/3$ if 8 or more independent trials give heads, where p denotes the probability of getting a head in any trial. Determine the sizes of type I and type II errors.*

Sol. This is clearly a binomial probability situation.

α=probability of rejecting H_0 when H_0 is true

=probability of rejecting H_0 when $p=1/2$

But H_0 is rejected when 8 or more trials give heads.

$$\alpha=\binom{10}{8}(1/2)^8(1/2)^2+\binom{10}{9}(1/2)^9(1/2)^1+\binom{10}{10}(1/2)^{10}(1/2)^0$$
$$=56/2^{10}.$$

β=probability of accepting H_0 when H_1 is true

=probability of accepting H_0 when $p=2/3$.

H_0 is accepted when the number of heads is less than 8

$$\beta=\binom{10}{0}(2/3)^0(1/3)^{10}+\binom{10}{1}(2/3)^1(1/3)^9+\dots$$
$$+\binom{10}{7}(2/3)^7(1/3)^3$$

$$=1-\left[\binom{10}{8}(2/3)^8(1/3)^2+\binom{10}{9}(2/3)^9(1/3)^1+\binom{10}{10}(2/3)^{10}(1/3)^0\right]$$

$$=1-17664/3^{10}.$$

10.21. The Critical Region. In the examples 10.2.1 and 10.2.2 the null hypothesis H_0 is tested on the basis of a sample of size one and 10 respectively. In Ex. 10.2.1, H_0 is rejected if the observed sample point (in this case the single observation) falls above 4000. The outcome set in this experiment of taking a single observation may be represented by the line $(0, \infty)$ since our observations are all positive because they are the consumption of electricity on various days.

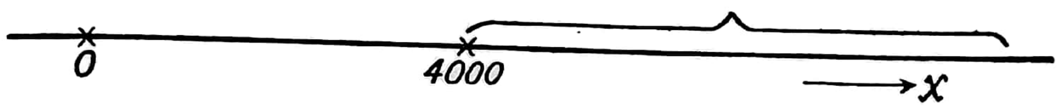


Fig. 10.2.

with regard to the test criterion in this example the outcome set (as this is a sampling problem we may very well call the outcome set the sample space) may be partitioned into two. H_0 is rejected if the observed sample point falls in one part and H_0 is accepted otherwise. This is illustrated in Fig. 10.2. The region of rejection of H_0 when H_0 is true or that region of the outcome set where H_0 is rejected if the sample point falls in that region, is called the critical region of the test. The probability that the sample point falls in the critical region is called the size of the critical region of the test. Evidently the size of the critical region is $\alpha=$ the probability of committing the type I error. If our test was based on a sample of size 2 in Ex. 10.2.1, then the outcome set or the sample space is the first quadrant in a two-dimensional space and a test criterion will enable us to separate our outcome set into two subsets. If the sample point falls in one subset, H_0 is rejected and H_0 is accepted otherwise. This is illustrated in Fig. 10.3 (*a*). In general if the outcome set is represented by a Venn diagram then the critical region is as shown in Fig. 10.3 (*b*).

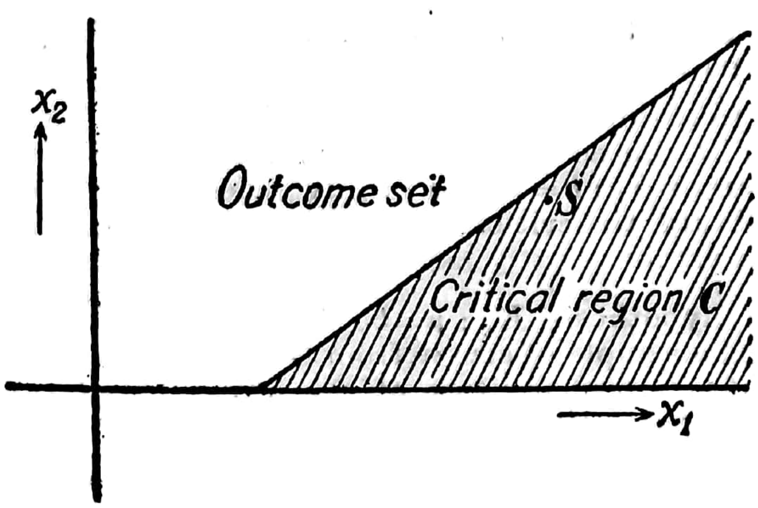


Fig. 10.3 (*a*)

The critical region is usually denoted by C. In the Ex. 10.2.1, the critical region C is the interval $(4000, \infty)$, This is shown in

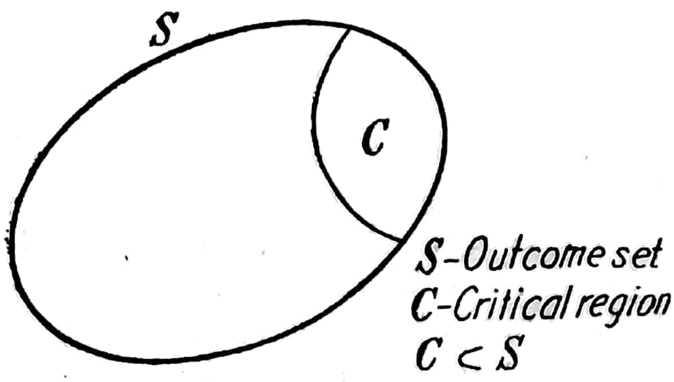


Fig. 10.3 (*b*)

Fig. 10.1. In Ex. 10.2.2 the outcome set may be given by a set of vectors of order 10 with the elements 1 and 0 where 1 denotes a head and 0 denotes a tail. The subset of vectors having 8 or more unities is the critical region C.

10.22. The best test for a simple hypothesis. Often the test criterion is to be determined by controlling α and β. If α and β can be simultaneously minimized it is desirable. But if α is minimized usually β becomes large and *vice versa*. This may be noticed from Fig. 10.1. Therefore a common practice is to select a test criterion which minimizes β for a fixed α. If there exists a test criterion which makes β a minimum for a fixed α then such a test is called the best test (best in the above sense) The existence of such a test when a simple hypothesis H_0 is tested against a simple alternative H_1, is given by a theorem due to J. Neyman and E.S. Pearson, which will be stated later.

Exercises

10.1. A test rejects the hypothesis that the probability of a son being born to a couple is 1/2, if the first two children are girls. Construct the outcome set and the critical region for this test. (Assume that we consider 10 children couples).

10.2. In a population $N(\mu, 1)$ the hypothesis that $\mu=2$ is rejected if a random sample of size 2 has a mean greater than 5, Obtain the sample space and the critical region for the test.

10 3. A coin is thrown 3 times. The hypothesis that the probability p of getting a head in any trial is 1/2 is rejected in favour of the hypothesis that $p=1/3$ if the three trials result in tails. Obtain the probabilities of the type I and the type II errors.

10.4. In a township the milk consumption of the families is assumed to be exponentially distributed with the parameter θ. The hypothesis $H_0 : \theta=5$ is rejected in favour of $H_1 : \theta=10$ if a family selected at random consumes 15 units or more. Obtain the critical region and the probabilities α and β of the type I and type II errors.

10.5. The hypothesis $H_0 : \mu=50$ is tested against the alternative $H_1 : \mu=60$ by using a sample of size n from the population $N(\mu, \sigma=5)$. How large should n be if the probabilities of the type I and type II errors are $\alpha=0{\cdot}025$ and $\beta=0.01$ respectively.

10.3. THE POWER OF A TEST

If the size of the critical region C is fixed (that is α is fixed) then that test which minimizes β may be called the best test in the case of a simple H_0 against a simple H_1. $1-\beta$=the probability of rejecting H_0 when it is not true. This is a correct decision andwe would like to have $1-\beta$ as close to one as possible. Hence $1-\beta$ is often called the power of the test whose critical region is C. A test criterion uniquely determines the critical region. This is seen in Section 10.21. Hence the power of a test may be called the power of a critical region. α is the size of the critical region or the probability that the sample point will fall in the critical region. We can have a number of tests for a null hypothesis H_0 with the same α but with different critical regions. This enables us to select that test for which the power $1-\beta$ is a maximum.

Ex. 10.3.1. *A box contains 10 marbles, out of which θ are red and the rest are green. We want to test the hypothesis $H_0 : \theta=5$ against the alternative $H_1 : \theta=4$. Determine the size of the critical regions and the power of the tests A and B.*

Test A. *Take two marbles at random with replacement and reject H_0 if both marbles are of the same colour.*

Test B. *Take two marbles at random with replacement and reject H_0 if both marbles are of different colours.*

Sol. $H_0 : \theta=5$

For test A, α=probability of rejecting H_0 when H_0 is true

=probability of rejecting H_0 when $\theta=5$

=probability of getting 2 marbles of the same colour when there are 5 red marbles.

$$=\binom{2}{2}(1/2)^2(1/2)^0+\binom{2}{2}(1/2)^0(1/2)^2$$

$$=2(1/2)^2=1/2.$$

For test B, α=probability of getting 2 marbles of different colours when there are 5 red marbles.

$$=\binom{2}{1}(1/2)^1(1/2)^1=1/2.$$

For test A, $1-\beta$=probability of rejecting H_0 when H_1 is true

=probability of getting 2 marbles of the same colour when there are 4 red marbles.

$$=\binom{2}{2}\left(\frac{4}{10}\right)^2\left(\frac{6}{10}\right)^0+\binom{2}{2}\left(\frac{4}{10}\right)^0\left(\frac{6}{10}\right)^2=13/25.$$

For test B, $1-\beta=\binom{2}{1}\left(\frac{4}{10}\right)^1\left(\frac{6}{10}\right)^1=12/25$

	α	$1-\beta$
Test A	1/2	13/25
Test B	1/2	12/25

Comments. Test A is more powerful than test B and hence B may be called non-admissible. In other words the critical region corresponding to A is more powerful than the critical region in B. This example is only for illustration. In a practical situation where we would like to test $H_0 : \theta=5$, we would take out all the marbles and count the number of red marbles. Sampling procedure is not needed everywhere.

When the alternative hypothesis is composite then $1-\beta$ may be evaluated for each value of the parameter specified by H_1. For example consider the hypothesis $H_0 : \theta=\theta_0$ against the alternative $H_1 : \theta\neq\theta_0$. For each θ not equal to θ_0 we can evaluate $1-\beta$ or the power of a test. If $1-\beta$ is plotted against θ, such a curve is called the power curve of a test. Let Fig. 10.4 give the power curves of three tests A, B and C with the same α. for testing H_0 ; $\theta=\theta_0$ against $H_1 : \theta\neq\theta_0$.

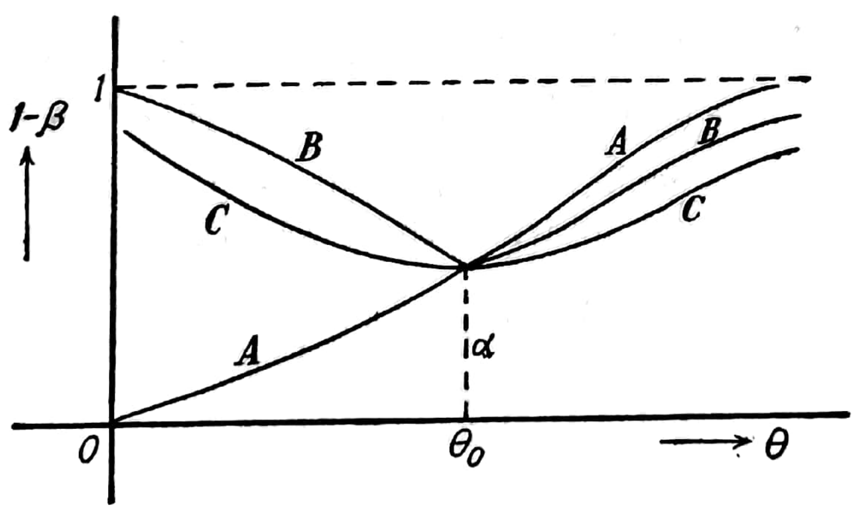


Fig. 10.4.

We would like to have $1-\beta$ as close to one as possible. Test A is more powerful than test B or C for $\theta>\theta_0$. Test B is uniformly more powerful than test C since the power curve for B lies closer to the line $1-\beta=1$ at every point than that of test C. Test A is less powerful than test B or C for $\theta<\theta_0$. Test C is said to be non-admissible compared to test B. If there exists a test which is uniformly more powerful than any other test it is called the uniformly most powerful or the best test. A method of obtaining the most powerful test whenever it exists, will be given in a later theorem.

Ex. 10.3.2. *Draw the power curve of the test in Ex. 10.2.1. for testing $H_0 : \theta=1000$ against the one sided alternative that $H_1 : \theta>1000$.*

Sol. In Ex. 10.2.1, the parent distribution is exponential and the test rejects the null hypothesis if an observed value is greater than or equal to 4000.

$$\alpha=\int_{4000}^{\infty} \frac{1}{1000} e^{-x/1000} dx=e^{-4}$$

The power of this test $=1-\beta=$probability of rejecting H_0 when H_1 is true

i.e.,
$$1-\beta=\int_{4000}^{\infty}\frac{1}{\theta} e^{-x/\theta} dx$$

where θ is any value >1000 (since H_1 is $\theta>\theta_0=1000$)

For $\theta=2000, 3000, 4000$ etc., $1-\beta$ will be e^{-2}, $e^{-4/3}$, e^{-1} etc. The power curve is obtained by plotting $1-\beta$ against θ where $\theta>1000$. This is given in Fig. 10.5.

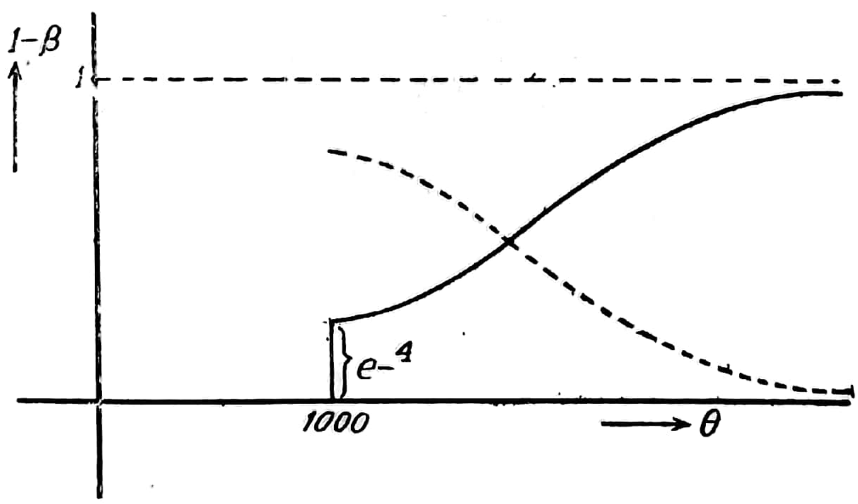


Fig. 10.5.

Comments. This is the power curve of a test which is based on a single observation. For different tests we can draw the corresponding power curves. Instead of the power curve if β=probability of accepting H_0 when H_1 is true, is plotted against θ then we get, what is usually called an operation characteristic curve (OC-curve). In industrial applications OC curve is more often used than the power curve in order to compare tests and to take decisions. The OC-curve for Ex. 10.3.2 is given in Fig. 10.5 as the dotted curve. It is obtained when β is plotted against θ.

Ex. 10.3.3. *A coin is thrown 6 times. The null hypothesis $H_0 : p=\frac{1}{2}$ is rejected if 5 or more trials result in heads. Obtain the*

power curve of this test if $H_0 : p=\frac{1}{2}$ is tested against $H_1 : p\neq\frac{1}{2}$, where p denotes the probability of getting a head in any trial.

Sol. α=probability of rejecting H_0 when H_0 is true

$$=\binom{6}{5}\left(\tfrac{1}{2}\right)^5\left(\tfrac{1}{2}\right)^1+\binom{6}{6}\left(\tfrac{1}{2}\right)^6\left(\tfrac{1}{2}\right)^0=7/64.$$

$1-\beta$ =probability of rejecting H_0 when H_1 is true.

$H_1 : p\neq 1/2$. Hence p may be any value between 0 and 1 but not equal to 1/2. Therefore,

$$1-\beta=\binom{6}{}(p)^5(1-p)^1+\binom{6}{6}p^6(1-p)^0$$

For various values of p, $1-\beta$ may be obtained from a binomial probability table. The power curve is given in Fig. 10.6.

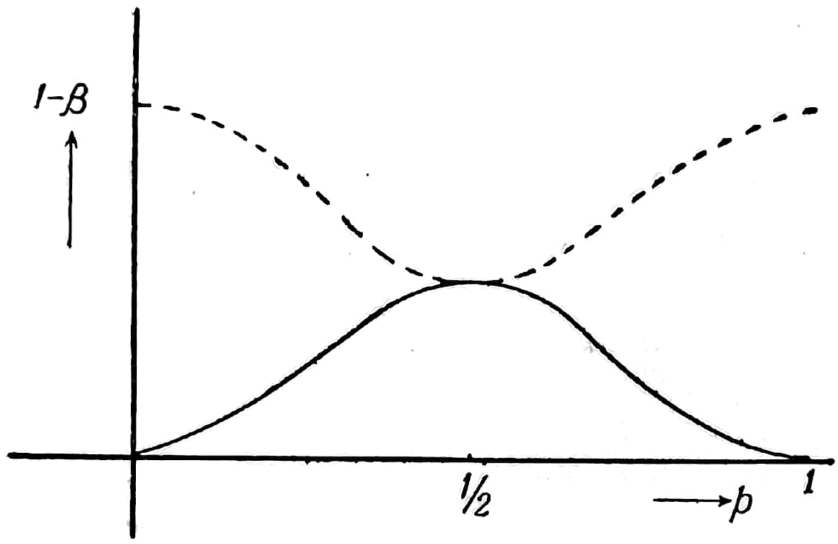


Fig. 10.6.

Comments. If β is plotted against p the OC-curve for the test is obtained. It is given by the dotted curve in Fig. 10.6.

So far we have been considering the various aspects of the theory of testing hypotheses when a test criterion is given. The following theorems will help us to obtain different test criteria.

Theorem 10.1. The Neyman-Pearson Lemma. Consider the problem of testing a simple hypothesis $H_0 : \theta=\theta_0$ against a simple alternative $H_1 : \theta=\theta_1$. Let $x_1, x_2, \ldots x_n$ be a random sample from the population $f(x, \theta)$ under consideration. The likelihood function under H_0 and H_1 are

$$L_0=\prod_{i=1}^{n} f(x_i, \theta_0) \qquad (10.6)$$

and

$$L_1=\prod_{i=1}^{n} f(x_i, \theta_1) \qquad (10.7)$$

respectively. For example if $f(x, \theta)=\frac{1}{\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2}}$

and if $H_0 : \mu=5$ and $H_1 ; \mu=10$ then

$$L_0 = \frac{1}{(\sqrt{2\pi})^n} e^{-\sum_{i=1}^{n} (x_i-5)^2/2}$$

and
$$L_1 = \frac{1}{(\sqrt{2\pi})^n} e^{-\sum_{i=1}^{n} (x_i-10)^2/2.}$$

We have defined the best test in a simple H_0 against a simple H_1 case, as that test with given α for which β is a minimum.

The Neyman-Pearson lemma says that if there exists a critical region C of size α and a constant k such that

$$\frac{L_0}{L_1} < k \text{ inside C} \tag{10.8}$$

and
$$\frac{L_0}{L_1} > k \text{ outside C.} \tag{10.9}$$

then C is the most powerful critical region for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$.

We have seen that in a problem of testing a simple H_0 against simple H_1 the selection of a critical region is equivalent to the selection of a test criterion. The inequalities (10.8) and (10.9) give the best test whenever it exists. Since $L_0/L_1 = k$ defines a set of measure zero, in the continuous case the inequalities (10.8) and (10.9) may be written as

$$\frac{L_0}{L_1} \leqslant k \text{ inside C} \tag{10.10}$$

and
$$\frac{L_0}{L_1} \geqslant k \text{ outside C.} \tag{10.11}$$

Proof. C is of size α. Let D be another critical region of the same size α. C and D are two regions in an n-dimensional space (since the sample is of size n). A symbolic representation of C and D is given in Fig. 10.7).

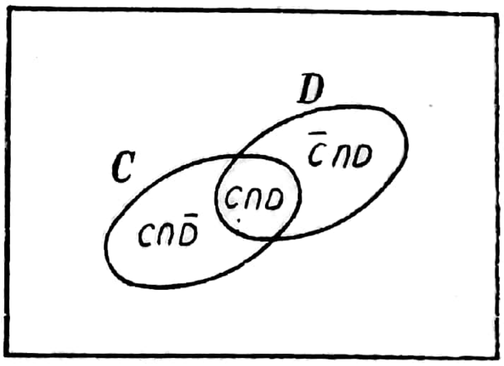


Fig. 10.7.

$$\int_C L_0\, dx = \int_D L_0\, dx = \alpha \text{ (by the definition of a critical region)} \tag{10.12}$$

where $dx = dx_1, dx_2 \ldots dx_n$ and the single integral stands for a multiple integral over the n-dimensional regions C and D

$$C = (C \cap \bar{D}) \cup (C \cap D) \text{ and } D = (\bar{C} \cap D) \cup (C \cap D) \tag{10.13}$$

and further $C \cap D$, $\bar{C} \cap D$, $C \cap \bar{D}$ are disjoint.

Hence equation (10.12) implies that

$$\int_{\bar{C} \cap D} L_0\, dx = \int_{C \cap \bar{D}} L_0\, dx. \tag{10.14}$$

$$\int_C L_1\, dx = \int_{C \cap D} L_1\, dx + \int_{C \cap \bar{D}} L_1\, dx > \int_{C \cap D} L_1\, dx + \int_{\bar{C} \cap D} \frac{L_0}{k} dx$$

$$\left(\text{since } L_1 > \frac{L_0}{k} \text{ inside C} \right) \tag{10.15}$$

$$\int_C L_1\, dx > \int_{C \cap D} L_1\, dx + \frac{1}{k} \int_{C \cap \bar{D}} L_0\, dx =$$

$$\int_{C \cap D} L_1\, dx + \frac{1}{k} \int_{\bar{C} \cap D} L_0\, dx \text{ [by equation (10.14)]} \tag{10.16}$$

But $\displaystyle\int_{\bar{C} \cap D} \frac{L_0}{k}\, dx > \int_{\bar{C} \cap D} L_1\, dx \left(\text{since } L_1 < \frac{L_0}{k} \text{ outside C} \right)$

$$\tag{10.17}$$

Equations (10.16) and (10.17) $\Rightarrow$

$$\int_C L_1\, dx > \int_{C \cap D} L_1\, dx + \int_{\bar{C} \cap D} L_1\, dx = \int_D L_1\, dx$$

$$\int_C L_1\, dx > \int_D L_1\, dx \tag{10.18}$$

i.e., the power of C is $>$ the power of D.

Hence β for C is $< \beta$ for D.

C is the most powerful critical region. The proof for a discrete case is similar and is left to the reader. The same theorem

can be used in some cases to obtain the best test when testing a simple H_0 against a composite H_1. This may be seen from the following examples. Whenever a sufficient statistic for the parameter exists it can be easily seen that the best critical region is a function of the value of the sufficient statistic. For this and other related topics such as unbiasedness, similar critical regions, most powerful unbiased tests etc., see the books listed in the Bibliography at the end of this chapter.

Ex. 10.3.4. *$x_1, x_2, \ldots, x_n$ is a random sample from a $N(\mu, \sigma)$ where σ is known. Obtain the best critical region of size α for testing $H_0 : \mu=\mu_0$ against $H_1 : \mu=\mu_1$ where μ_0 and μ_1 are specified values.*

Sol. This is a case of a simple H_0 against a simple H_1.

Theorem 10.1 can be used to obtain the best test.

$$L_0=\frac{1}{(\sigma\sqrt{2\pi})^n}\, e^{-\sum_{i=1}^{n}(x_i-\mu_0)^2/2\sigma^2} \tag{10.19}$$

$$L_1=\frac{1}{(\sigma\sqrt{2\pi})^n}\, e^{-\sum_{i=1}^{n}(x_i-\mu_1)^2/2\sigma^2} \tag{10.20}$$

$$\frac{L_0}{L_1}=e^{\left[\frac{n}{2}\left(\mu_1^2-\mu_0^2\right)+\left(\mu_0-\mu_1\right)\Sigma x_i\right]/\sigma^2}$$

(obtained by simplification

(10.20)

According to the theorem 10.1 the best critical region C is given by

$$\frac{L_0}{L_1}\leqslant k \text{ inside C}$$

and $$\frac{L_0}{L_1}\geqslant k \text{ outside C when } k \text{ is a constant.}$$

$$\frac{L_0}{L_1}\leqslant k \Rightarrow e^{\left[\frac{n}{2}\left(\mu_1^2-\mu_0^2\right)+(\mu_0-\mu_1)\,\Sigma x_i\right]/\sigma^2}\leqslant k. \tag{10.22}$$

Taking logarithms and simplifying the inequality we get

$(\mu_0-\mu_1)\,\bar{x}\leqslant k'$ where k' is a constant

and $$\bar{x}=\Sigma\, x_i/n. \tag{10.23}$$

Case I. Let $\mu_1>\mu_0$

Then $\mu_0-\mu_1<0$ and division of (10.23) by $\mu_0-\mu_1$ gives $\bar{x}\geqslant K$ where K is a constant. (10.24)

The best critical region is given by the inequality (10.24). K is easily determined since the size of the critical region is α.

i.e. $$\int_K^{\infty} f(\bar{x})\, d\bar{x} \text{ (when } \mu=\mu_0)=\alpha \qquad (10.25)$$

But we know that $K=\mu_0+Z_\alpha \dfrac{\sigma}{\sqrt{n}}$ where Z_α is given in Fig. 10.8 which gives the distribution of $\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$; N(0, 1).

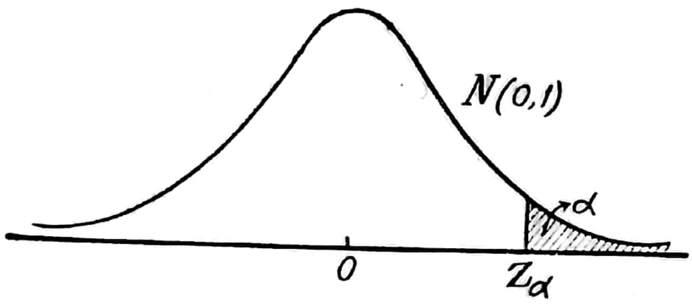


Fig. 10.8.

Hence the test can be stated as follows. When the observed sample mean $\bar{x} \geqslant \mu_0+Z_\alpha \dfrac{\sigma}{\sqrt{n}}$ reject H_0, otherwise accept H_0.

i.e., when $$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \geqslant Z_\alpha \text{ reject } H_0 \qquad (10.26)$$

where σ, μ_0, n are known and Z_α is obtained from a normal probability table.

Case II. Let $\mu_1<\mu_0$ then $\mu_0-\mu_1>0$.

Division of (10.23) by $\mu_0-\mu_1$ gives

$$\bar{x} \leqslant K' \text{ inside C} \qquad (10.27)$$

where K′ is a constant which can be easily determined.

$$\int_{-\infty}^{K'} f(\bar{x})d\bar{x} \text{ (when } \mu=\mu_0) = \alpha \qquad (10.28)$$

i.e., $$K'=\mu_0-Z_\alpha \cdot \frac{\sigma}{\sqrt{n}}.$$

This is illustrated in Fig. 10.9.

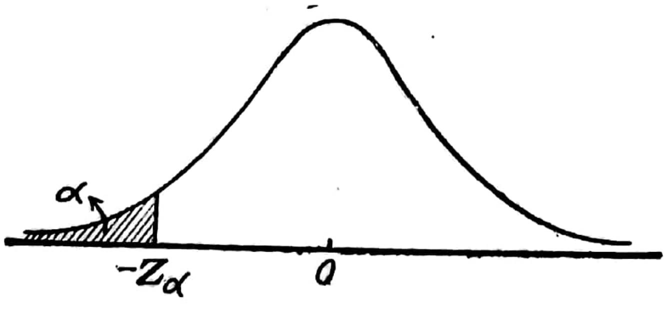


Fig. 10.9.

The best test in this case may be given as follows.

When $\bar{x} \leqslant \mu_0 - Z_\alpha \dfrac{\sigma}{\sqrt{n}}$ reject H_0

i.e., when $\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \leqslant -Z_\alpha$ reject H_0, (10.29)

otherwise accept H_0, where μ_0, σ, n are known and Z_α is determined from a normal probability table.

In these two cases we get one sided tests. Usually one-sided alternatives, that is, alternatives of the type $\theta > \theta_0$ or $\theta < \theta_0$ lead to one sided tests. Here we have seen that our success in finding the best test depends on getting a statistic whose distribution is independent of the parameters. For example, in the above cases we know that the distribution of $(\bar{X}-\mu_0)/(\sigma/\sqrt{n})$ is independent of μ and σ and hence Z_α is easily obtained and hence the inequalities (10.26) and (10.29) are obtained.

Ex. 10.3.5. *The yield of a particular crop is assumed to be distributed as a $N(\mu, \sigma=2)$. A random sample of 9 test plots give an average yield of 25 units. Test the hypothesis that the true average yield $\mu=20$ against the alternative that $\mu>20$ at the 5% level.*

Sol. $X : N(\mu, \sigma)$; $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} : N(0, 1)$

The Neyman-Pearson lemma leads to the test criterion that if $\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \geqslant Z_{0\cdot 05}$ reject H_0, otherwise accept H_0.

Here $\bar{x}=25$, $\mu_0=20$, $\sigma=2$, $n=9$ and $Z_{0\cdot 05}=1.64$; $Z_{0\cdot 05}$ is obtained from a normal probability table.

$$\int_{1.64}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = 0.05$$

$$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{(25-20)}{2/3} = 7.5 > 1.64$$

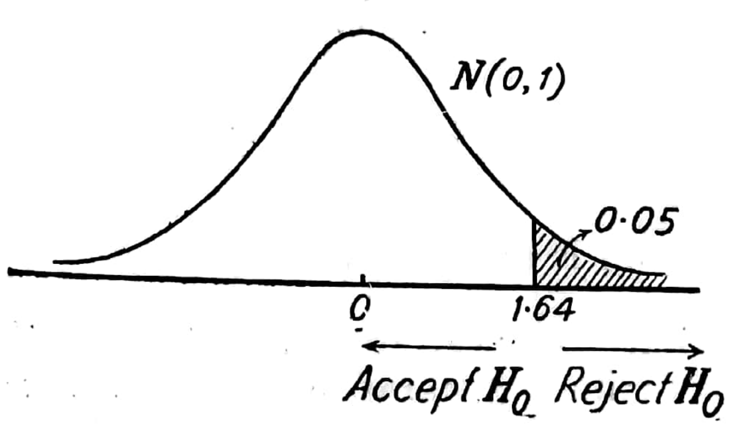


Fig. 10.10.

Hence $H_0 : \mu=20$ is rejected in favour of $H_1 : \mu>20$.

Ex. 10.3.6. *The weights of the students in a particular grade are assumed to be a $N(\mu, \sigma=5)$. A random sample of 25 students give a total weight equal to 1250 units. Test the hypothesis that $\mu=52$ against the alternative $\mu<52$ at the 1% level.*

Sol. Here $\bar{x}=\dfrac{1250}{25}=50$, $\mu_0=52$, $\sigma=5$, and $n=25$

$$X : N(\mu, \sigma)$$

Hence a convenient statistic is

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} : N(0, 1)$$

The Neyman-Pearson lemma leads to the test criterion that if $\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \leqslant -Z_{0 \cdot 01}$ reject H_0 : otherwise accept H_0.

Here $Z_{0 \cdot 01}=2.32$ (obtained from a normal probability table).

$$\int_{-\infty}^{-2.32} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt=0.01$$

$$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}=\frac{(50-52)}{5/5}=-2>-2.32$$

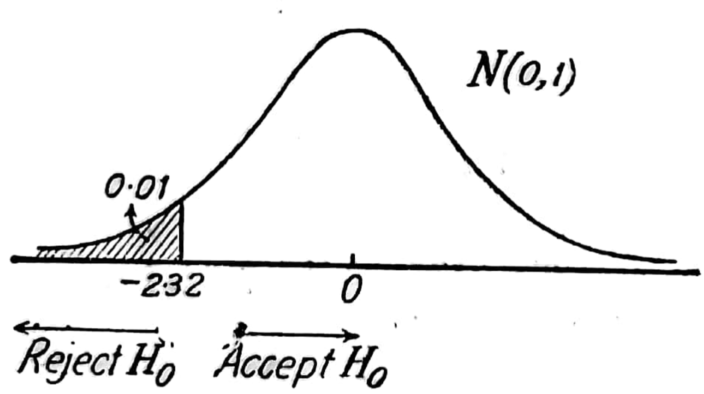


Fig. 10.11.

Hence $H_0 : \mu=52$ is accepted.

Theorem 10.2. The Likelihood Ratio Test. This test can be considered to be a generalization of the Neyman-Pearson lemma for testing a simple H_0 against a simple H_1.

Consider a problem of testing a composite hypothesis or a simple H_0 against a composite H_1 ; *i.e.*, hypothesis of the type

$$\begin{cases} H_0 : \theta=\theta_0 \\ H_1 : \theta\neq\theta_0 \end{cases} ; \begin{cases} H_0 : \theta>\theta_0 \\ H_1 : \theta<\theta_0 \end{cases} \begin{cases} H_0 : \theta^{(1)}=\theta_0^{(1)} & \text{(some of the parameters are specified)} \\ H_1 : \theta^{(1)}\neq\theta_0^{(1)} & \text{etc.} \end{cases}$$

Let max L_0 = the maximum value of the likelihood function L under the null hypothesis H_0. This is obtained by substituting the maximum likelihood estimates of the parameter in L under H_0. If H_0 specifies all the parameters in L then max. $L_0=L_0$ itself. Let max. L be the maximum of L with respect to the parameters and

$$\lambda=\frac{\text{max. } L_0}{\text{max. } L}. \tag{10.30}$$

λ is called the likelihood ratio statistic or the λ-criterion. The likelihood ratio test says that a uniformly most powerful critical region C for testing H_0, simple or composite, is usually obtained by the inequality

$$\lambda\leqslant k \text{ inside C} \tag{10.31}$$

where the constant k is usually determined by the inequality that the probability of the type I error $\leqslant\alpha$. (10.31) does not always give a uniformly most powerful test. Sometimes it leads to a non-admissible test. Evidently $0<\lambda\leqslant1$ and intuitively λ is a reasonable test statistic for H_0.

In the case of a composite hypothesis we can formulate the hypothesis in the following general terms.

$$H_0 : \theta \in \omega$$

$$H_1 : \theta \in \bar{\omega} \cap \Omega \tag{10.32}$$

where $\omega \subset \Omega$ and Ω is the parameter space. Ω is the space generated by all possible values of the parameters. For example in an exponential distribution

$$f(x, \theta)=\begin{cases} \frac{1}{\theta}\, e^{-x/\theta} \text{ for } x>0 \text{ and } \theta>0 \\ 0 \text{ elsewhere} \end{cases}$$

the parameter $\theta>0$ and hence Ω is the open interval 0 to ∞

$$\Omega=(0, \infty).$$

In a normal distribution $N(\mu, \sigma)$, $-\infty<\mu<\infty$ and $\sigma>0$.

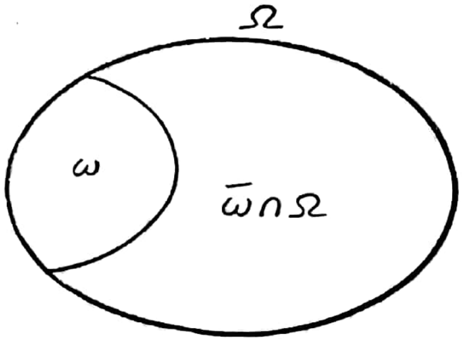


Fig. 10.12.

Therefore Ω is the upper half plane if μ is measured on the x-axis and σ on the y-axis. If there are k parameters in a distribution Ω is a subspace of a k-dimensional space. In general if Ω is denoted by a Venn diagram then the hypothesis H_0 and H_1 are illustrated in Fig. 10.12.

$$H_0 : \theta \varepsilon \omega \ ; \ H_1 : \theta \varepsilon \overline{\omega} \cap \Omega$$

where θ stands for all the parameters in the distribution under consideration.

Ex. 10.3.7. *Obtain the critical region by using the likelihood ratio test for testing H_0 : $\mu=\mu_0$ against H_1 : $\mu\neq\mu_0$ in a $N(\mu, \sigma)$ where σ is known. The probability of the type I error is given to be $\leqslant\alpha$ and a random sample $x_1, x_2 \ldots\ldots x_n$ is observed.*

Sol. $$L=\frac{1}{(\sqrt{2\pi}\ \sigma)^n}\, e^{-\sum\limits_{i=1}^{n}(x_i-\mu_0)^2/2\sigma^2}$$

Here σ is known and H_0 : $\mu=\mu_0$. Therefore

$$\text{max. } L_0=L_0=\frac{1}{(\sqrt{2\pi\sigma})^n}\, e^{-\sum\limits_{i=1}^{n}(x_i-\mu_0)^2/2\sigma^2}. \qquad (10.33)$$

The maximum likelihood estimate of μ is $\bar{x}=\Sigma x_i/n$

$$\text{max. } L=\frac{1}{(\sqrt{2\pi\sigma})^n}\, e^{-\sum\limits_{i=1}^{n}(x_i-\bar{x})^2/2\sigma^2} \qquad (10.34)$$

$$\lambda=\frac{\text{max } L_0}{\text{max. } L}=\frac{e^{-\sum\limits_{i=1}^{n}(x_i-\mu_0)^2/2\sigma^2}}{e^{-\Sigma(x_i-\bar{x})^2/2\sigma^2}} \qquad (10.35)$$

$$=e^{-n(\bar{x}-\mu_0)^2/2\sigma^2} \qquad (10.36)$$

$$\lambda \leqslant k \Rightarrow$$

$$e^{-n(\bar{x}-\mu_0)^2/2\sigma^2} \leqslant k \Rightarrow \qquad (10.37)$$

$(\bar{x}-\mu_0)^2 \geqslant k'$ where k' is a constant or (10.38)

$$|\,\bar{x}-\mu_0\,| \geqslant k''. \qquad (10.39)$$

The best critical region is determined by the inequality (10.39) and is obtained by using the result that the probability of the type I error $\leqslant\alpha$.

$$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} : N(0, 1)$$

We know that $P\left\{\left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| \geqslant Z_{\alpha/2}\right\}=\alpha$

where $Z_{\alpha/2}$ is illustrated in Fig. 9.13.

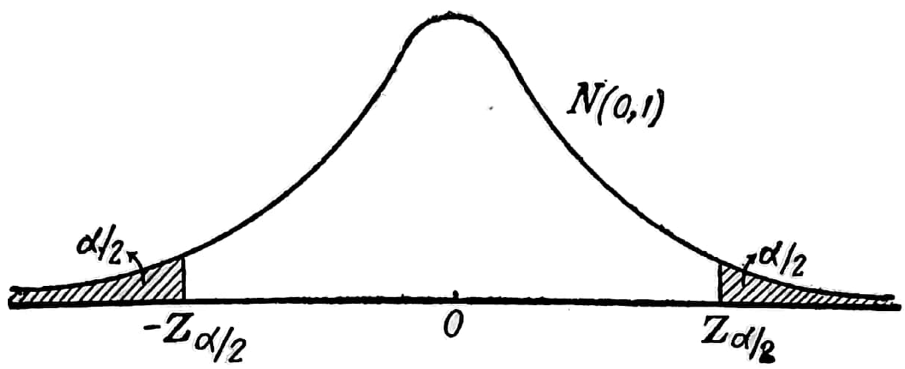


Fig. 10.13

The critical region is determined by the inequality

$$\left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| \geqslant Z_{\alpha/2} \tag{10.40}$$

or

$$|\bar{x}-\mu_0| \geqslant Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{10.41}$$

Ex. 10.3.8. *The incomes per week of the citizens in a township are assumed to have a distribution $N(\mu, \sigma=5)$. A random sample of 25 people shows an average income of \$90 per week. Test the hypothesis that the true average income $\mu=\$100$ against the alternative $\mu\neq\$100$. Use $\alpha=0.05$.*

Sol. $\bar{x}=90$, $n=25$, $\mu_0=100$, $\sigma=5$ and $Z_{\alpha/2}=1.96$

$$X : N(\mu, \sigma)$$

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} : N(0, 1)$$

By the likelihood ratio test if $\left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| \geqslant 1.96$, H_0 is rejected.

$$\left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right|=\left|\frac{90-100}{(5/5)}\right|=10>1.96$$

Hence H_0 is rejected

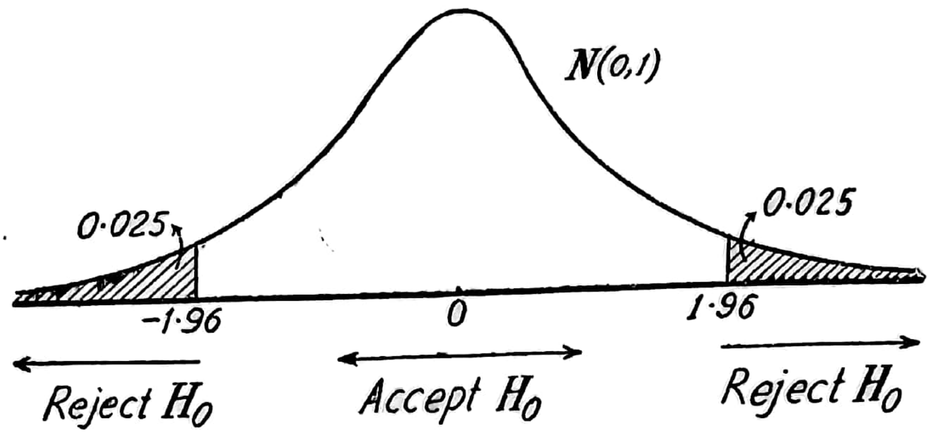


Fig. 10.14

Comments. It is seen from Ex. 10.3.4, Ex. 10.3.5 and Ex. 10.3.7 that usually one sided alternatives lead to one sided tests and a two sided alternative leads to a two sided test. This is not true in general.

Theorem 10.3. We shall state the following theorem without proof. Under very general conditions the distribution of 2 $\log_e \lambda$ approaches a χ^2 distribution with its degrees of freedom equal to the number of parameters that are determined by the hypothesis H_0, when n is sufficiently large, where λ is the likelihood ratio criterion given by equation (10.30).

For example in testing H_0 $\mu=\mu_0$ and $\sigma=\sigma_0$ against $H_1 : \mu\neq\mu_0$ and $\sigma\neq\sigma_0$, in a $N(\mu, \sigma)$ based on a random sample of size n, $-2\log_e \lambda$ is approximately a χ^2 with 2 degrees of freedom since two parameters are specified by H_0. When the parent population is a $N(\mu, \sigma)$ where σ is known we can show that for any n, $-2 \log_e \lambda$ for testing $H_0 : \mu=\mu_0$ against $H_1 : \mu\neq\mu_0$, is a χ^2 with 1 degree of freedom.

In this case

$$-2 \log_e \lambda = \frac{n}{\sigma^2}(\bar{x}-\mu_0)^2 \qquad \text{(shown in Ex. 10.3.7)}$$

$$= \left(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right)^2 \qquad (10.42)$$

But when $X : N(\mu, \sigma)$, $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is a (N(0, 1) and therefore

$\left(\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right)^2$ is a χ^2 with one degree of freedom.

Theorem 10.3, does not specify the parent population. According to this theorem for large n, whatever may be the parent population the null hypothesis H_0 is rejected when,

$$-2 \log \lambda \geqslant \chi^2_{\alpha,r} \qquad (10.43)$$

where $\chi^2_{\alpha,r}$ denotes the point such that $P\left\{\chi^2_r \geqslant \chi^2_{\alpha,r}\right\}=\alpha$; r stands for the number of parameters specified by H_0 (the number of degrees of freedom of the χ^2) and α stands for the level of significance.

Ex. 10.3.9. *$X_1, X_2, \ldots X_k$ are k independent normal variables with means $\mu_1, \mu_2, \ldots, \mu_k$ and with variance $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$. Random. samples of sizes $n_1, n_2, \ldots, n_k$ are taken from these k populations. Test the hypothesis $H_0 : \sigma_1=\sigma_2=\ldots=\sigma_k$ at the 100 α% level.*

Sol. Let $x_{i1}, x_{i2}, \ldots x_{in_i}$ be the observed random sample from the i^{th} population. $i=1, 2, \ldots k$.

The joint density function of the random sample for the i^{th} population

$$=\frac{1}{\sigma_i^{n}\,(\sqrt{2\pi}\,)^{n_i}}\, e^{-\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2/2\sigma_i^2} \tag{10.44}$$

The likelihood function L is therefore given by

$$L=\prod_{i=1}^{k}\frac{1}{\sigma_i^{n_i}(\sqrt{2\pi})^{n_i}}\, e^{-\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2/2\sigma_i^2} \tag{10.45}$$

$$=\frac{1}{(\sqrt{2\pi})^{n}\left(\prod_{i=1}^{k}\sigma_i^{n_i}\right)}\, e^{-\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{i}-\mu_i)^2/2\sigma_i^2}$$

where $n=n_1+n_2+\ldots+n_k$ (10.46)

The maximum likelihood estimates of $\mu_1, \mu_2, \ldots, \mu_k$ and $\sigma_1, \sigma_2, \ldots, \sigma_k$ are obtained by maximizing L with respect to these parameters. The estimates are easily seen to be

$$\hat{\mu}_i=\sum_{j=1}^{n_i} x_{ij}/n_i=\bar{x}_i \text{ and } \hat{\sigma}_i^2=\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2/n_i=s_i^2 \text{ (say)} \tag{10.47}$$

The maximum value of L is obtained by substituting these estimates in L

$$\max.\ L=\frac{1}{(\sqrt{2\pi})^{n}\left(\prod_{i=1}^{k} s_i^{n_i}\right)}\, e^{-n/2}$$

$$H_0 : \sigma_1 = \sigma_2 = \ldots = \sigma_k = \sigma \text{ (say)} \tag{10.48}$$

H_0 specifies $k-1$ parameters. If all the variances are assumed to have specific values then H_0 specifies k parameters. Here the variances are assumed to be equal and hence only $k-1$ independent parameters are specified.

The likelihood function L under H_0 is

$$L_0 = \frac{1}{\sigma^n (\sqrt{2\pi})^n} \, e^{-\Sigma\,\Sigma (x_{ij} - \mu_i)^2 / 2\sigma^2} \tag{10.49}$$

Since $\prod_{i=1}^{k} \sigma_i^{n_i} = \sigma^{n_1 + \ldots + n_k} = \sigma^n$ when $\sigma_1 = \ldots = \sigma_k = \sigma$.

The maximum likelihood estimates of the parameters under H_0 are obtained by maximizing L_0 with respect to $\mu_1, \mu_2, \ldots \mu_k$ and σ. The estimates are easily seen to be

$$\hat{\mu}_1 = \sum_{j=1}^{n_i} (x_{ij})/n_i = \bar{x}_i \text{ and } \hat{\sigma}^2 = \frac{n_1 s_1^2 + \ldots + n_k s_k^2}{n} \tag{10.50}$$

for $i = 1, 2, 3 \ldots k$.

max. L_0 is obtained by substituting these estimates in L_0.

$$\max L_0 = \frac{1}{(\sqrt{2\pi})^n \sigma} e^{-n/2} \tag{10 51}$$

$$\lambda = \frac{\max L_0}{\max L} = \frac{s_2^{n_1} \; s_2^{n_2} \ldots s_k^{n_k}}{\left(\dfrac{n_1 s_1^2 + \ldots + n_k s_k^2}{n} \right)^{n/2}}. \tag{10.52}$$

Since H_0 specifies $k-1$ parameters, $-2 \log_e \lambda$ may be assumed to have a χ^2 distribution with $k-1$ degrees of freedom when n is sufficiently large. The hypothesis H_0 is rejected when

$-2 \log \lambda \geqslant \chi^2_{\alpha, k-1}$ when n is large, where λ is given by (10 52)

and $\chi^2_{\alpha, k-1}$ is the tabulated value of a χ^2_{k-1} at the 100 α% level.

Comments. The problem of testing equality of variances is of great practical importance. For small values of n_i a modified statistic which is a modification of $-2 \log \lambda$ and which has an approximate χ^2 distribution with $k-1$ degrees of freedom is often suggested.

$$-2 \log \mu/[1+\frac{1}{3(k-1)}(\sum_{i=1}^{k} \frac{1}{n_i-1}-\frac{1}{n-k})] : \chi^2_{k-1} \text{approximately}$$

$$\text{where } \mu = \prod_{i=1}^{k} \left(\frac{n_i\, s_i^2}{n_i-1}\right)^{\frac{n_i-1}{2}} \Bigg/ \left(\frac{\Sigma\, n_i\, s_i^2}{\Sigma(n_i-1)}\right)^{\frac{\Sigma(n_i-1)}{2}} \qquad (10.54)$$

Exercises

10.6. By using the Neyman-Pearson Lemma obtain the best test for testing the following hypothesis : (*a*) $H_0 : \sigma_0=\sigma_0$, $H_1 : \sigma>\sigma_0$ in a $N(0, \sigma)$, (*b*) $H_0 : \theta=\theta_0$, $H_1 : \theta>\theta_0$ in an exponential population, (*c*) $H_0 : p=p_0$, $H_1 : p>p_0$ in a Binomial population with the parameters n and p (Assume that the probability of the type I error is α and a sample of size n is taken).

10.7. Obtain the best test, if it exists, for testing the hypothesis $H_0 : \theta=5$ against $H_1 : \theta=6$, in the population $f(x, \theta)=1/\theta$ for $0<x<\theta$. Assume that a single observation is taken and $\alpha=0.05$.

10.8. In the problem 10.7 if the critical region is given as $x \geqslant 4.3$ obtain α and β.

10.9. For testing the hypothesis $H_0 : \theta=5$ in the population $f(x, \theta)=1/\theta$ for $0<x<\theta$, against the alternative $H_1 : \theta\neq5$ draw the power curve if the test is based on a single observation and if the critical region is $x \geqslant 4.5$ or $x \leqslant 0.5$.

10.10. In a shipment of 10 articles θ are defective. The hypothesis $H_0 : \theta=5$ is rejected if two articles drawn at random without replacement are either both good or both defective, otherwise the hypothesis is accepted. Obtain β if H_1 is $\theta=0, 1, 2, 3, 4, 6, 7, 8, 9, 10$ and plot the power curve and the OC-curve for this test.

10.11. All the 12 students in a class are classified into two groups according to their aptitude for Mathematics. θ of them are interested in Mathematics and the others are not interested in Mathematics. The hypothesis $H_0 : \theta=6$ is tested against the alternative $H_1 : \theta\neq6$ by the following tests (1) Two students are selected at random with replacement and H_0 is rejected if both are either in one group or in the other group ; (2) Two students are selected at random with replacement and H_0 is rejected if they belong to different groups. Is one of these critical regions non-admissible ? Draw the power curves for the two critical regions.

10.12. For the hypothesis $H_0 : \sigma=2$ is tested against $H_1 : \sigma\neq2$ in a $N(0, \sigma)$. Draw the power curve and the OC curve for the critical region if $\alpha=0.05$, assuming that the test is based on a random sample of size 9.

10.13. $N(\mu_1, 1), N(\mu_2, 1), \ldots, N(\mu_k, 1)$ are k independent populations. Obtain the likelihood ratio test for testing the hypothesis $H_0 : \mu_1=\ldots=\mu_k$ against the alternative that all the μ's are not equal, assuming that the test is based on random samples of sizes $n_1, \ldots, n_k$ respectively and the size of the critical region is α.

10.14. A random sample of size n is taken from a $N(\mu, \sigma)$. For testing the hypothesis $H_0 : \mu=\mu_0$, show that the likelihood ratio criterion λ is a function of a student t. Obtain the distribution of $\lambda^{2/n}$.

10.15. Obtain the likelihood ratio criterion for testing $H_0 : \mu_1=\ldots=\mu_k$, $\sigma_1=\ldots=\sigma_k$ where $N(\mu_1, \sigma_k), \ldots, N(\mu_k, \sigma_k)$ are independent normal populations. Assume that random samples of sizes $n_1, \ldots, n_k$ are taken from these populations.

10.16. Obtain the likelihood ratio criterion for testing $H_0 : \rho=0$ where ρ is the correlation coefficient in a bivariate normal population.

10.17. If a hypothesis $H_0 : \theta<\theta_0$ against $H_1 : \theta>\theta_0$ is tested about the parameter θ of an exponential distribution what are ω and Ω ?

10.4. TESTS CONCERNING MEANS

Consider a normal population $N(\mu, \sigma)$ where σ is known. The one sided and two sided tests concerning μ have already been discussed in section 10.3. When σ is unknown, for testing,

$H_0 : \mu=\mu_0$ against the alternative

(1) $H_1 : \mu<\mu_0$, (2) $\mu>\mu_0$, (3) $\mu\neq\mu_0$

a student t statistic can be conveniently used,

$$\frac{\overline{X}-\mu}{S'/\sqrt{n}} : t_{n-1} \tag{10.55}$$

that is, the statistic $(\overline{X}-\mu)/(S'/\sqrt{n})$ is a student t with $n-1$ degrees of freedom, where $S'^2=\Sigma(X_i-\overline{X})^2/(n-1)$ is an unbiased estimator of σ^2. The Neyman—Pearson lemma leads to the test criteria

$$\begin{cases} H_0 : \mu=\mu_0 \\ H_1 : \mu<\mu_0 \end{cases} \quad \text{If } \frac{\bar{x}-\mu_0}{(s'/\sqrt{n})}\leqslant -t_{\alpha,\, n-1} \text{ reject } H_0 \tag{10.56}$$

$$\begin{cases} H_0 : \mu=\mu_0 \\ H_1 : \mu>\mu_0 \end{cases} \quad \text{If } \frac{\bar{x}-\mu_0}{(s'/\sqrt{n})}\geqslant t_{\alpha,\, n-1} \text{ reject } H_0 \tag{10.57}$$

$$\begin{cases} H_0 : \mu=\mu_0 \\ H_1 : \mu\neq\mu_0 \end{cases} \quad \text{If } \left|\frac{\bar{x}-\mu_0}{(s'/\sqrt{n})}\right|\geqslant t_{\alpha/2,n-1} \text{ reject } H_0 \tag{10.58}$$

where $t_{\alpha,\, n-1}$ and $t_{\alpha/2,\, n-1}$ are the values of a student t with $n-1$ degrees of freedom such that

$$\int_{t_{\alpha,\, n-1}}^{\infty} f(t)\, dt=\alpha \text{ and } \int_{t_{\alpha/2,\, n-1}}^{\infty} f(t)\, dt=\alpha/2 \text{ respectively} \tag{10.59}$$

and $f(t)$ is the density function of a student t with $n-1$ degrees of freedom. These tests are illustrated in Fig. 10.15

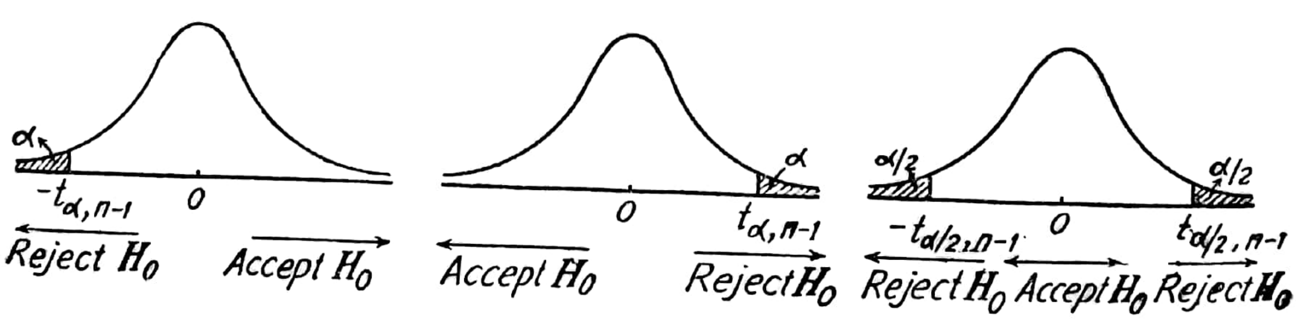


Fig. 10.15.

$\begin{cases} H_0 : \mu=\mu_0 \\ H_1 : \mu<\mu_0 \end{cases}$ $\begin{cases} H_0 : \mu=\mu_0 \\ H_1 : \mu>\mu_0 \end{cases}$ $\begin{cases} H_0 : \mu=\mu_0 \\ H_1 : \mu\neq\mu_0 \end{cases}$

When the sample size n is large (>30) then

$(\overline{X}-\mu)/(S'/\sqrt{n})$ is approximately normally distributed. Therefore when n is large, instead of a student t distribution, a normal distribution can be used.

Ex. 10.4.1. *A random sample of size 9 from a $N(\mu, \sigma)$ has mean 20 and variance 16. Test the hypothesis $H_0 : \mu=25$ against the alternative $\mu\neq25$ at the 5% level.*

Sol. $n=9$, $\bar{x}=20$, $s^2=\Sigma(x_i-\bar{x})^2/n=16$ and $\mu_0=25$.

$$s'^2=\frac{n}{n-1}\,s^2=\frac{9}{8}.\,16=18$$

$$|(\bar{x}-\mu_0)/(s'/\sqrt{n})|=|(20-25)/(18/3)|=5/6.$$

The tabulated value of a student t with $n-1=8$ degrees of freedom at the 5% level, that is, $t_{0.025,\,8}=2.31$ (obtained from a student t table)

i.e., $$\int_{2.31}^{\infty} f(t)\,dt=0.025$$

where $f(t)$ is the density function of a student t with 8 degrees of freedom. In this problem the alternative H_1 is $\mu\neq25$. Hence H_0 is rejected if $|(\bar{x}-\mu_0)/(s'/\sqrt{n})| \geqslant 2.31$.

But the observed value of the student t statistic is $5/6<2.31$ Hence the hypothesis is accepted.

Comments. If the sample size was >30 we could have based our test on a standardized normal variable,

$$\frac{\overline{X}-\mu_0}{S'/\sqrt{n}} : N(0, 1) \text{ approximately} \tag{10.60}$$

Testing of a simple hypothesis may be explained as follows. If our assumption, that the sample is a random sample from a $N(\mu, \sigma)$, is correct then under the hypothesis $H_0 : \mu=\mu_0$ the statistic $(\overline{X}-\mu_0)/(S'/\sqrt{n})$ is a student t with $n-1$ degrees of freedom. An observed value of this statistic falls between -2.31 and 2.31 with a probability equal to 0.95. In this problem we observe a value of this statistic. The value is 5/6 and which lies in between -2.31 and 2.31. If our observed value falls below -2.31 or above 2.31 the probability for such an event is less than 0.05. If our hypothesis is correct an improbable event has happened.

Hence we will be compelled to reject our hypothesis. Even if the parent population is not normal we can get a normal approximation for large samples by using the central limit theorem. Tests can be based on this approximation.

Exercises

10.18. A random sample of size 9 from a $N(\mu, \sigma=2)$ has a mean 50. Test the hypothesis, (1) $H_0: \mu=52$ against $H_1: \mu<52$; (2) $H_0: \mu=52$ against $H_1: \mu\neq 52$, at the 5% level.

10.19. The time taken by a particle to move from one fixed point to another fixed point is distributed as a $N(\mu, \sigma=0.1)$. 16 independent trials of this experiment give the average time equal to 10 units. Test the hypothesis that the expected duration of travel in any trial is 9 units against the alternative that it is more, at the 1% level.

10.20. The lifetime of television picture tubes produced by a particular factory is assumed to have an exponential distribution. One tube selected at random had a life time of 950 hrs. Test the factory's claim that the expected life time is 1000 hrs. against the alternative that it is less, at the 1% level.

10.21. Suppose that the energy used by a person for walking a unit distance is given by $e=(1/10)\ w+(1/4)\ A$ where e, w, A denote, energy, weight, and age respectively. An experiment conducted on a random sample of 16 twenty year old people from a certain city yields the following results : $\Sigma e_i=32$, $\Sigma e_i^2=72$ where e_i denotes the energy used by the ith person. Assuming that e's are distributed as a $N(\mu, \sigma)$ test the hypothesis that the expected weight of the 20 year olds in the city is 120 against the alternative that it is not, at the 1% level.

10.22. A feeding experiment conducted on 100 experimental animals shows an average increase in weight of 10 lbs with a standard deviation of 2 lbs. Test the hypothesis that the expected increase is 12 lbs against the alternative that it is not, at the 5% level.

[**Hint :** Use a normal approximation].

10.41. Tests Concerning Differences between Means. There are many problems where we are interested in testing the difference between two population means. Suppose that we want to test the following hypothesis. (1) The yield of a particular variety of wheat exceeds that of another variety by 10 units ; (2) Drug A is as effective as drug B in curing a disease ; (3) Detergent A is more powerful than detergent B ; (4) The speed of a computing machine A is more than that of the machine B ; (5) The average I.Q of boys is higher than that of girls etc. All these problems can be considered to be cases of testing the difference between two population means. In this section we will consider a few simple cases of testing problems when the populations are normal.

Suppose, we have random samples of sizes n_1 and n_2 from two independent populations $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$, where σ_1 and σ_2 are known. Suppose that we would like to test the following hypothtses.

(a) $H_0 : \mu_1-\mu_2=\delta$ (b) $H_0 : \mu_1-\mu_2=\delta$ (c) $H_0 : \mu_1-\mu_2=\delta$

$H_1 : \mu_1-\mu_2<\delta$ $H_1 : \mu_1-\mu_2>\delta$ $H_1 : \mu_1-\mu_2\neq\delta$

(10.61)

We know that

$$Z=\frac{(\overline{X}_1-\overline{X}_2)-\delta}{\left[\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right]^{\frac{1}{2}}} \quad \text{is a } N(0, 1). \tag{10.62}$$

It can be easily seen that the likelihood ratio technique will give the following test criteria for testing the hypotheses in (10.61).

The null hypothesis H_0 is rejected if

$$z\leqslant -z_\alpha \text{ in } (a)\ ;\ z\geqslant z_\alpha \text{ in } (b)\ ;\ |z| \geqslant z_{\alpha/2} \text{ in } (c) \tag{10.63}$$

where z is an observed value of Z, z_α and $z_{\alpha/2}$ are such that

$$\int_{z_\alpha}^{\infty}\frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt=\alpha \text{ and } \int_{z_{\alpha/2}}^{\infty}\frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt=\alpha/2. \tag{10.64}$$

If σ_1 and σ_2 are unknown then the statistic

$$Z=\frac{(\overline{X}_1-\overline{X}_2)-(\delta)}{\left[\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right]^{\frac{1}{2}}} \quad : N(0, 1) \text{ approximately} \tag{10.65}$$

when n_1 and n_2 are large, where $\overline{X}_1$ and $\overline{X}_2$ are the sample means and S_1^2 and S_2^2 are the sample variances.

$$S_1^2 = \sum_{i=1}^{n_1} (X_{1i}-\overline{X}_1)^2/n_1 \text{ and } S_2^2 = \sum_{i=1}^{n_2} (X_{2i}-\overline{X}_2)^2/n_2. \tag{10.66}$$

For the various hypotheses in (10.61) tests may be constructed based on the statistic in (10.65). These tests are easily seen to be the same as the tests in (10.63).

In the normal populations if $\sigma_1=\sigma_2=\sigma$ or if the populations are $N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$ and if σ is known then the likelihood ratio technique leads to the test statistic

$$Z=\frac{(\overline{X}_1-\overline{X}_2)-\delta}{\sigma\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{\frac{1}{2}}} : N(0, 1). \tag{10.67}$$

For testing the various hypotheses in (10.61) the test criteria are similar to the ones given in (10.63).

If $\sigma_1=\sigma_2=\sigma$ where σ is unknown then the likelihood ratio technique will lead to the following statistic t which is a student t with n_1+n_2-2 degeees of freedom

$$t_{n_1+n_2-2}=\frac{(\overline{X}_1-\overline{X}_2)-\delta}{S\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{\frac{1}{2}}} \tag{10.68}$$

where $S^2=\left(n_1 S_1^2+n_2 S_2^2\right) \Big/ \left(n_1+n_2-2\right)$ and S_1^2 and S_2^2 are the sample variances. In this case the various test criteria for testing the hypotheses in (10.61) are as follows. Reject the hypothesis H_0 if

$$t\leqslant -t_{\alpha,\, n_1+n_2-2} \text{ in } (a)\ ;\ t\geqslant t_{\alpha,\, n_1+n_2-2} \text{ in } (b)\ ;$$

$$|\, t\, | \geqslant t_{\alpha/2,\, n_1+n_2-2} \text{ in } (c) \tag{10.69}$$

where t is an observed value of $t_{n_1+n_2-2}$ in (10.68)

and $t_{\alpha,\, n_1+n_2-2}$ and $t_{\alpha/2,\, n_1+n_2-2}$ are such that

$$\int_{t_{\alpha,\, n_1+n_2-2}}^{\infty} f(t)\, dt=\alpha \quad\text{and}\quad \int_{t_{\alpha/2,\, n_1+n_2-2}}^{\infty} f(t)\, dt=\alpha/2 \tag{10.70}$$

where $f(t)$ is the density function of a student t with n_1+n_2-2 degrees of freedom. When $n_1+n_2-2\geqslant 30$ the student t statistic in (10.68) approximates N(0, 1) and therefore the tests in this case can be based on a N(0, 1) variable.

Ex. 10.41.1. *Two random samples of sizes 10 and 12 of I.Q's of men and women, have means 101 and 98 respectively. Assuming that the I.Q's are independently normally distributed as $N(\mu_1, \sigma_1=4)$ and $N(\mu_2, \sigma_2=3)$, test the hypothesis $H_0 : \mu_1=\mu_2$ against the alternative $H_1 : \mu_1\neq\mu_2$ at the 5% level.*

Sol. $\bar{x}_1=101$, $\bar{x}_2=98$, $n_1=10$, $n_2=12$, $\sigma_1=4$, $\sigma_2=3$.

and
$$\alpha=0.05$$

$$X_1 : N(\mu_1, \sigma_1=4)$$

$$X_2 : N(\mu_2, \sigma_2=3)$$

$$Z=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\left[\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right]^{\frac{1}{2}}} : N(0, 1)$$

where $\mu_1-\mu_2=0$ under H_0. (10.71)

Hence H_0 is rejected if $|z| \geqslant z_{0.025} = 1.96$.

$z_{\alpha/2} = z_{0.025}$ is obtained from a normal probability table.

When H_0 is true

$$|Z| = \frac{|101-98-0|}{\sqrt{\frac{16}{10}+\frac{9}{12}}} = 1.89 < 1.96$$

The hypothesis is accepted.

Comments. If we want to test the hypothesis that $\mu_1>\mu_2$ we usually formulate the hypothesis $H_0 : \mu_1=\mu_2$ and test H_0 against $H_1 : \mu_1>\mu_2$. This is why H_0 is usually called the null hypothesis. We take the hypothesis H_0 in the form $\mu_1=\mu_2$ only for convenience. In general we may define the null hypothesis as that hypothesis whose false rejection is considered to be a type I error.

Ex. 10.41.2. *The average yields of 10 and 20 test plots of two varieties of wheat are 30 and 40 with standard deviations 4 and 6 respectively. If the yields of the two varieties of wheat are assumed to be independently and normally distributed as $N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$ respectively, test whether there is any significant difference between the average yields of the two varieties, at the 5% level.*

Sol. $\bar{x}_1=30$, $\bar{x}_2=40$, $n_1=10$, $n_2=20$, $\sigma_1=\sigma_2=\sigma$

$s_1=4$, $s_2=6$ and $\alpha=0.05$

$X_1 : N(\mu_1, \sigma)$

$X_2 : N(\mu_2, \sigma)$

Let $H_0 : \mu_1=\mu_2$ (There is no difference between the average yields)

$H_1 : \mu_1 \neq \mu_2$

When H_0 is true $\mu_1-\mu_2=0$

$$\frac{(\overline{X}_1-\overline{X}_2)-0}{\sigma\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{\frac{1}{2}}} : N(0, 1) \qquad (10.72)$$

But σ is unknown. σ may be estimated by S where

$$S^2=\frac{n_1 S_1^2+n_2 S_2^2}{n_1+n_2-2}$$

and

$$\frac{(\bar{X}_1-\bar{X}_2)-0}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} : t_{n_1+n_2-2}$$

$H_0 : \mu_1=\mu_2$ is rejected if

$$\left|\frac{\bar{x}_1-\bar{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right| \geqslant t_{\alpha/2,\, n_1+n_2-2} = t_{0.025,28} = 2.048.$$

$t_{0.025,28}=2.048$ is obtained from a student t table.

$$\left|\frac{\bar{x}_1-\bar{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right| = \left|\frac{30-40}{5\ 6\sqrt{\frac{1}{10}+\frac{1}{20}}}\right| \geqslant 2.048$$

The hypothesis is to be rejected.

Comments. Here $H_0 : \mu_1=\mu_2$ is rejected. Hence if our assumptions of normality etc., are correct, the observed difference between $\bar{x}_1$ and $\bar{x}_2$ cannot be attributed to chance alone. We can be on the safe side in assuming such an inference in this case. If we know further that the only other variation in the observations is the difference between the effects of the two varieties of wheat we can possibly say that the varieties may be considered to be different as far as their yields are concerned. Tests of significance will be discussed in the last chapter. Instead of two varieties if we had k varieties and if we wanted to test the hypothesis $H_0 : \mu_1=\mu_2=\ldots\ldots=\mu_k$ against H_1 : not all μ's are equal, then a test criterion could be constructed by using the likelihood ratio technique. A general method of dealing with such problems, called the analysis of variance technique, will be discussed in the last chapter.

The problem of testing $H_0 : \mu_1=\mu_2$ when $\sigma_1\neq\sigma_2$ where σ_1 and σ_2 are unknown and when n_1 and n_2 are not large, does not come under any of the cases discussed so far. This problem is called the Behrens-Fisher problem. This will not be discussed here. For this and related topics, see Kenney, J.F. and E.S. Keeping Mathematics of Statistics, Part 2, Van Nostrand Co., New York, 1951 and other references in the bibliography at the end of this chapter. The reader should take particular care in the philosophy of 'testing hypotheses'. Since a hypothesis is different from a fact whenever we accept a hypothesis, it does not mean that the hypothesis is true. Acceptance of a hypothesis that

$\theta=\theta_0$ does not mean that $\theta=\theta_0$. θ may or may not be equal to, θ_0. In other words we are not making a statement, 'therefore $\theta=\theta_0$' or we are not proving anything.

Exercises

10.23. Two diets are compared by conducting an experiment on two sets of 40 and 50 experimental animals. The average increase in weights due to the diets A and B are 10 lbs and 12 lbs with standard deviations 2 lbs and 3 lbs respectively. Check the claim that diet B increases the weight by 3 lbs more than that of diet A on the average.

10.24. Two methods of teaching are compared by teaching two classes of 20 and 25 students by two teachers whose I.Q's are 110 and 120 respectively. The marks obtained by the students are assumed to have the distributions $N(T_1+I_1/10, \sigma)$ and $N(T_2+I_2/10, \sigma)$ respectively, where T and I denote the effect of teaching and the I.Q. of the teachers respectively. Test the hypothesis that the two methods are equally effective, at the 5% level. The average marks of the two classes are 65 and 70 with standard deviations 2 and 3 respectively.

10.25 The money spent by the customers at two stores selling the same goods, is assumed to have the distributions $N(3B_1+2Q_1, \sigma_1=2)$ and $N(3B_2+2Q_2, \sigma_2=3)$ respectively, where B and Q denote the beauty of the salesgirl in the department and the quality of the goods respectively. Check the claim that $B_1>B_2$ if $Q_1=Q_2$ and if the following observations are made. 15 customers spent a total of \$200 at the first shop and 20 customers spent a total of \$260 at the second shop.

10.26. In a laboratory for weight reduction, a random sample of 10 women, shows an average weight of 150 lbs with a standard deviation of 5 lbs before undergoing the treatment and an average weight of 140 lbs with a standard deviation of 4 lbs after undergoing the treatment. The weight distributions before and after the treatment are assumed to be $N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$ respectively. Check the claim that the special treatment is effective in reducing weight, at the 5% level. Assume that the populations are independent.

10.27. A random sample of 300 persons in city A shows an average annual income of \$10,000 with a standard deviation of \$300. A random sample of 400 citizens in city B shows an average income of \$9,000 with a standard deviation of \$250. Check whether there is any significant difference between the incomes of the citizens in cities A and B on the average.

[**Hint.** Use a normal approximation and choose a suitable level of significance.]

10.28. For testing the hypothesis $H_0: \mu_1-\mu_2=20$ against the alternative $H_1: \mu_1-\mu_2=25$, in the populations $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$, obtain n if random samples of the same size n are taken and if $\sigma_1=4$, $\sigma_2=5$, $\alpha=0.05$ and $\beta=0.05$.

10.5. TESTS CONCERNING VARIANCES

In section 10.41, we have seen that for testing $H_0: \mu_1=\mu_2$ when $\sigma_1=\sigma_2$ and when $n_1+n_2-2 \geqslant 30$ a student t statistic can be used. As a pre-requisite of this test we have the problem of testing the equality of variances. There are many practical situations where we would like to test the equality of variabilities. In this section we will consider only the special case of test concerning the variances when the populations are normal.

Consider the single sample problem of testing

$H_0 : \sigma^2 = \sigma_0^2$ against the alternative,

(a) $H_1 : \sigma^2 > \sigma_0^2$; (b) $H_1 : \sigma^2 < \sigma_0^2$,

(c) $H : \sigma^2 \neq \sigma_0^2$. (10.73)

That is, we have a random sample of size n from a $N(\mu, \sigma)$ and we want to test the hypothesis $\sigma^2 = \sigma_0^2$ where σ_0 is a specified quantity.

We know that $\dfrac{nS^2}{\sigma^2} : \chi^2_{n-1}$ (10.74)

i.e., nS^2/σ^2 is a χ^2 with $n-1$ degrees of freedom, where S^2 is the sample variance. The likelihood ratio technique will lead to the following tests. Reject H_0 if the observed value of nS^2/σ^2 when $\sigma^2 = \sigma_0^2$ is such that

(a) $\geqslant \chi^2_{\alpha,\, n-1}$; (b) $\leqslant \chi^2_{1-\alpha,\, n-1}$;

(c) $\geqslant \chi^2_{\alpha/2,\, n-1}$ or $\leqslant \chi^2_{1-\alpha/2,\, n-1}$

where the points

$\chi^2_{\alpha,\, n-1}$, $\chi^2_{\alpha/2,\, n-1}$, $\chi^2_{1-\alpha,\, n-1}$ and $\chi^2_{1-\alpha/2,\, n-1}$

are illustrated in Fig. 10.16.

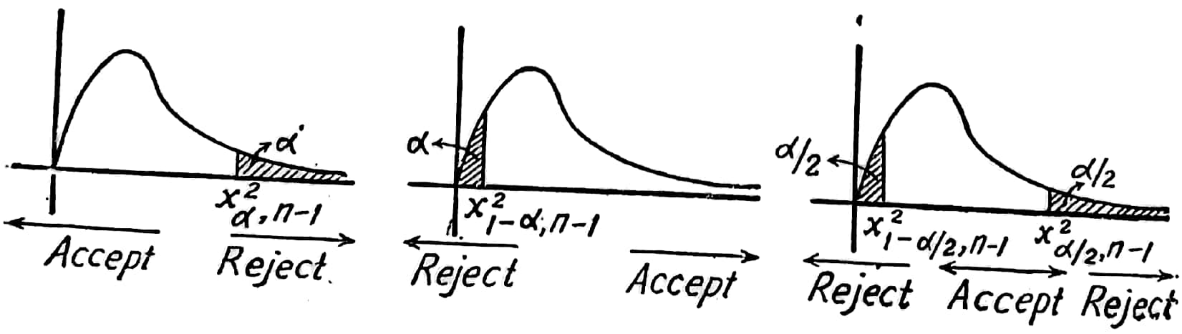


Fig. 10.16.

$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$ $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases}$ $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$

Fig. 10.16 gives the distribution of a χ^2 with $n-1$ degrees of freedom.

Ex. 10.5.1. *A random sample of 20 electric bulbs produced according to a special process, have an average life of 1000 hrs with a standard deviation of 10 hrs. Assuming that the life time of these bulbs has a distribution $N(\mu, \sigma)$ test the hypothesis $H_0: \sigma=9$ against $H_1: \sigma>9$ at the 1% level.*

Sol. $n_1=20$, $s=10$, $\sigma_0=9$ and $\alpha=0.01$

$$nS^2/\sigma^2 : \chi^2_{n-1} \tag{10.75}$$

This is a convenient statistic and when $\sigma=\sigma_0=9$

$$nS^2/\sigma_0{}^2=\frac{20(100)}{9}=222.2.$$

This is the observed value of a χ^2 with $n-1=19$ degrees of freedom. $H_1: \sigma>9$ and therefore H_0 is rejected if the observed value of the χ^2_{19} is greater than or equal to

$$\chi^2_{\alpha, n-1} = \chi^2_{0.01,19} = 36.191$$

(obtained from a χ^2 table)

$$222.2>36.191.$$

Hence the hypothesis $H_0: \sigma=9$ is rejected in favour of $H_1: \sigma>9$.

Comments. When there are a number of populations $N(\mu_1, \sigma_1)$, $N(\mu_2, \sigma_2)$,......, $N(\mu_k, \sigma_k)$ for testing the hypothesis $H_0: \sigma_1=\sigma_2=\ldots=\sigma_k$ against the alternative H_1: not all σ's are equal. Theorem 10.3 will give an appropriate test criterion.

Let there be two random samples of sizes n_1 and n_2 from two independent normal populations $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$ respectively. Consider the following tests.

$$H_0: \sigma_1^2=\sigma_2^2 \text{ against the alternative } H_1.$$

(a) $H_1: \sigma_1^2 > \sigma_2^2$; (b) $H_1: \sigma_1^2 < \sigma_2^2$

(c) $\sigma_1^2 \neq \sigma_2^2$. (10.76)

The likelihood ratio technique will lead to the following tests.

Reject the hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ if

$$(a) \quad \frac{n_1 s_1^2 / (n_1-1)}{n_2 s_2^2 / (n_2-1)} > F_{\alpha, n_1-1, n_2-1} \tag{10.77}$$

$$(b) \quad \frac{n_2 s_2^2 / (n_2-1)}{n_1 s_1^2 / (n_1-1)} > F_{\alpha, n_2-1, n_1-1}$$

$$(c) \quad \left.\begin{array}{ll} \dfrac{n_1 s_1^2 / (n_1-1)}{n_2 s_2^2 / (n_2-1)} > F_{\alpha/2, n_1-1, n_2-1} & \text{if } s_1^2 > s_2^2 \\[2ex] \dfrac{n_2 s_2^2 / (n_2-1)}{n_1 s_1^2 / (n_1-1)} > F_{\alpha/2, n_2-1, n_1-1} & \text{if } s_2^2 > s_1^2 \end{array}\right] \tag{10.79}$$

Here all the tests are based on the statistic

$$F_{n_1-1, n_2-1} = \frac{n_1 S_1^2 / \sigma_1^2 (n_1-1)}{n_2 S_2^2 / \sigma_2^2 (n_2-1)}$$

$$= \frac{n_1 S_1^2 / (n_1-1)}{n_2 S_2^2 / (n_2 \quad 1)} \quad \text{when } \sigma_1^2 = \sigma_2^2 \tag{10.80}$$

and all the test criteria are based on the right tail area of an F distribution. This is achieved because of the property that if

$$X : F_{n_1-1, n_2-1} \text{ then } \frac{1}{X} : F_{n_2-1, n_1-1}.$$

F_{α, n_1-1, n_2-1} is the point such that

$$\int_{F_{\alpha, n_1-1, n_2-1}}^{\infty} f(F)dF = \alpha \tag{10.81}$$

where $f(F)$ is the density function of an F- with n_1-1 and n_2-1 degrees of freedom.

Ex. 10.5.2. *The diameters of two random samples each of size 10, of bullets produced by two machines, have standard deviations $s_1=0.01$ and $s_2=0.015$. Assuming that the diameters have independent distributions $N(\mu, \sigma_1)$ and $N(\mu, \sigma_2)$, test the hypothesis that the two machines are equally good by testing $H_0 : \sigma_1=\sigma_2$ against $H_1 : \sigma_1\neq\sigma_2$.*

Sol. $s_1=0.01$, $s_2=0.015$, $n_1=n_2=10$, and let $\alpha=0.02$

$$\frac{n_1 S_1^2 \Big/ (n_1-1)}{n_2 S_2^2 \Big/ (n_2-1)} : F_{n_1-1,\, n_2-1} \qquad (10.82)$$

Since $s_2>s_1$ and $H_1 : \sigma_1\neq\sigma_2$ the hypothesis $H_0 : \sigma_1=\sigma_2$ is rejected when

$$\frac{n_2 s_2^2 \Big/ (n_2-1)}{n_1 s_1^2 \Big/ (n_1-1)} \geqslant F_{\alpha/2,\, n_2-1,\, n_1-1} = F_{0.01, 9, 9}$$

[see (10.79)]

Here $$\frac{n_2 s_2^2 \Big/ (n_2-1)}{n_1 s_1^2 \Big/ (n_1-1)} = \frac{(0.015)^2}{(0.01)^2} = 2.25$$

But $F_{0.01,9,9}=5.35$ (obtained from an F-table) *i.e.*, $2.25<5.35$ and therefore $H_0 : \sigma_1=\sigma_2$ is accepted. The machines may be considered to be equally good.

Exercises

10.29. The thickness of metal plates produced by a machine, is assumed to have a distribution $N(\mu, \sigma)$. A random sample of 20 plates shows a standard deviation of 0.01 units. Test the hypothesis $H_0 : \sigma=0.02$ against (1) $H_1 : \sigma<0.02$, (2) $\sigma>0.02$, (3) $\sigma\neq0.02$, at the 5% level.

10.30. The weight of paper bags is assumed to have a distribution $N(\mu, \sigma)$. A random sample of 40 bags shows a standard deviation of 0.05 units. Test the hypothesis $H_0 : \sigma=0.03$ against $H_1 : \sigma\neq0.03$, at the 5% level.

[**Hint.** $\sqrt{2\chi^2}-\sqrt{2n-1}$ is approximately a standardized normal when $n>30$].

10.31. The following data give the time spent by random samples of 10 boys and 8 girls for solving a problem. Assuming that the samples may

be considered to be from $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$ respectively, test the hypothesis $\sigma_1^2 = \sigma_2^2$ against the alternative $\sigma_1^2 \neq \sigma_2^2$, at the 10% level.

Girls. 20, 22, 18, 15, 16, 18, 20, 18.

Boys. 18, 20, 20, 22, 16, 14, 12, 8, 10, 8.

10.6. TESTS CONCERNING PROPORTIONS

So far we have been considering the problem of testing hypotheses concerning the parameters of a continuous population. In this section we shall briefly discuss the problem of testing of hypotheses regarding the parameters of a discrete distribution. For convenience and simplicity we shall consider only a binomial probability situation ; that is, testing whether, a coin is unbiased, a particular drug reduces the mortality rate by 15%, the true proportion of defective items in a large shipment is 10% etc. The general problem can be formulated as follows :

$$H_0 : p = p_0$$

$$H_1 : (a)\ p > p_0,\ (b)\ p < p_0,\ (c)\ p \neq p_0 \tag{10.83}$$

where p_0 is a specified value of the probability of a success in any trial of a binomial probability situation. Our tests will be based on the number of successes observed in a sample of n trials. The likelihood ratio technique will give the following tests Reject the hypothesis H_0 if

$$(a)\ x \geqslant x_0 \text{ where } \sum_{x=x_0}^{n} f(x, n, p_0) \leqslant \alpha \tag{10.84}$$

$$(b)\ x \leqslant x_1 \text{ where } \sum_{x=0}^{x_1} f(x, n, p_0) \leqslant \alpha \tag{10.85}$$

$$(c)\ x \geqslant x_3 \text{ or } \leqslant x_2 \text{ where } \sum_{x=0}^{x_2} f(x, n, p_0) \leqslant \alpha/2 \text{ and}$$

$$\sum_{x=x_3}^{n} f(x, n, p_0) \leqslant \alpha/2 \tag{10.86}$$

where $f(x, n, p_0)$ is the binomial probability function with the parameters n and p_0, α is the probability of the type I error and x_0, x_1, x_2 x_3 are the nearest integers which satisfy the various inequalities in (10 84), (10.85) and (10.86).

Ex. 10.6.1. *Out of 20 babies born in a given hospital 12 are girls. Assuming a binomial probability situation, test the hypothesis that the probability of birth of a baby girl is 1/2 against the alternative that it is greater than 1/2, at the 5% level.*

Sol. Let p be the probability of birth of a baby girl

$$p_0=1/2,\ n=20,\ x=12, \text{ and } \alpha=0.05$$

$H_1: p>p_0$ and therefore H_0 is rejected if $x \geqslant x_0$ where x_0 is such that

$$\sum_{x=x_0}^{n} f(x, n, p_0) \leqslant 0.05$$

From a binomial probability table for $n=20$ and $p=1/2$

it is seen that $\displaystyle\sum_{x=15}^{20} \binom{20}{x} (1/2)^x (1/2)^{20-x} \leqslant 0.05$

Therefore $x_0=15$ and the observed value is 12.

Hence the hypothesis is accepted at the 5% level.

Comments. When the number of trials n is sufficiently large, more specifically, when $np>5$ and $nq>5$, a normal approximation to the binomial is valid. In this case our test may be based on the normal variable

$$Z=\frac{X-np}{\sqrt{npq}} : N(0, 1) \tag{10 87}$$

where x denotes the number of successes in n trials and $q=1-p$. The test criteria for testing the various hypotheses in (10.83) are therefore, reject H_0 when

(a) $z \geqslant z_\alpha$; (b) $z \leqslant -z_\alpha$; (c) $|z| \geqslant z_{\alpha/2}$

where z and $z_{\alpha/2}$ are such that, z is the observed value of Z and

$$\int_{z_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt=\alpha \text{ and } \int_{z_{\alpha/2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt=\alpha/2. \tag{10.88}$$

The two sample problem (the problem of testing the difference between p_1 and p_2 in two independent binomial situations with the parameters n_1, p_1 and n_2, p_2 respectively) can be treated in a similar fashion. Tests concerning k proportions is discussed in the next chapter.

In all the problems discussed in this chapter we made decisions based on a sample of pre-assigned size n. This may be an unnecessary restriction. We might be able to make the same decisions by taking fewer observations. Another drawback of the methods discussed so far, is that in certain problems three may be more than two possible choices. We considered only two

choices, namely, either accept the hypothesis or reject the hypothesis. The process of decision making when there are a number of choices available is sometimes called the multiple decision problem. Instead of taking a sample of pre-assigned size and testing a hypothesis, we may decide to take additional observations only after considering the information available so far ; this method is called a sequential testing procedure. For example we may start with a sample of size m. The null hypothesis is tested. Suppose that the choices are (1) accept H_0, (2) reject H_0, (3) continue sampling. If our decision is to continue sampling we will take one more observation and test H_0. Depending upon the result of this test either sampling is continued or H_0 is accepted or rejected. The likelihood ratio test, modified to suit the three choices, can be used in the sequential procedure. For multiple decision problems and sequential procedures see the bibliography at the end of this chapter.

Exercises

10.32. In a binomial probability situation of 20 trials, if p is the probability of a success and if the hypothesis $H_0 : p=0.3$ is tested against the alternative $p=0.2$, 0.5, 0.7 obtain the probability of the type II error if $\alpha=0.05$, in each case.

10.33. In a binomial probability situation of 15 trials obtain a test for testing the hypothesis $H_0 : p=0.60$ against the alternative $H_1 : p>0.60$, if $\alpha=0.05$.

10.34 Out of 20 patients who are given a particular injection 18 survived. Will you reject the hypothesis that the survival rate is 85% in favour of the hypothesis that it is more, at the 5% level ?

10.35. A random sample of 1000 persons in a country shows that 550 favoured a particular legislature. Test the hypothesis that more than 50% of the people favour the legislature, at the 1% level.

10.36. A public opinion survey conducted on random samples of 400 women and 600 men, shows that 200 women and 325 men are in favour of erecting a monument at a particular place. Test the hypothesis that the true proportions, are the same, against the alternative that they are not the same, at the 4% level.

10.37. (2, 5) is a random sample of size 2 from a Poisson population with parameter λ. Obtain a test for testing the hypothesis $H_0 : \lambda=4$ against the alternative $H_1 : \lambda\neq4$, if $\alpha=0.05$.

10.38. What must be the size of the sample if $\alpha=0.05$, $\beta=0.08$ for testing $H_0 : \lambda=1$ against $H_1 : \lambda=1.5$ where λ is the parameter of a Poisson population.

10.39. The following data give the yields of independent random test plots of three varieties of wheat :

A 42, 42, 44, 47, 43, 46, 42, 44, 46, 40, 38, 42, 45, 47.

B 42, 41, 40, 41, 38, 40, 42, 45, 46, 40, 39, 41, 44, 48.

C 38, 42, 44, 46, 40, 44, 48, 50, 52, 51, 49, 46, 48, 50.

Assuming that these samples may be considered to be from $N(\mu_1, \sigma_1)$, $N(\mu_2, \sigma_2)$, $N(\mu_3, \sigma_3)$ respectively, test the following hypotheses at the 1% level.

(1) $\mu_1=\mu_2=\mu_3$, $\sigma_1=\sigma_2=\sigma_3=2$ (given) ;

(2) $\sigma_1=\sigma_2=\sigma_3$;

(3) $\mu_1=\mu_2=\mu_3$, $\sigma_1=\sigma_2=\sigma_3=\sigma$ and σ is unknown,

10.7. SUMMARY

A summary of the simple tests discussed in this chapter, is given in the following table. For convenience only two-sided tests are given except for problem 1. In problem 1, all the one-sided and two-sided alternatives are discussed. For other situations which are not discussed in the following table the likelihood ratio technique will help us to obtain appropriate test criteria. Single sample tests are based on a random sample of size n and two sample tests are based on random samples of sizes n_1 and n_2. The usual notations for the sample mean, the sample variance, etc., are used. All the tests are assumed to have a critical region of size α. In two sample problems the populations are assumed to be independent. In problems 7 and 8 of the table, $\hat{\sigma}^2$ is an unbiased estimator of σ^2

$$\hat{\sigma}^2=\left(n_1\, S_1^2 + n_2 S_2^2\right)\Big/(n_1+n_2-2)$$

$$S_1^2 = \sum_{i=1}^{n_1} (X_{1i}-\overline{X}_1)^2/n_1 \;;$$

$$S_2^2 = \sum_{i=1}^{n_2} (X_{2i}-\overline{X}_2)^2/n_2 \;;$$

$$S'^2= \sum_{i=1}^{n} (X_i-\overline{X})^2/(n-1).$$

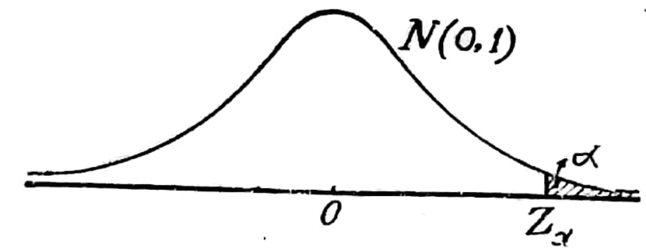
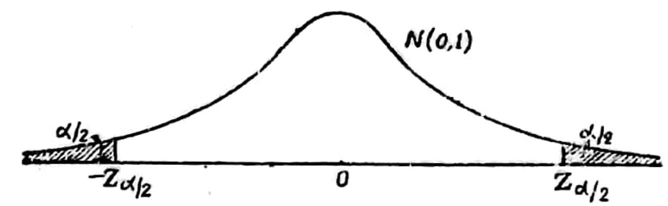
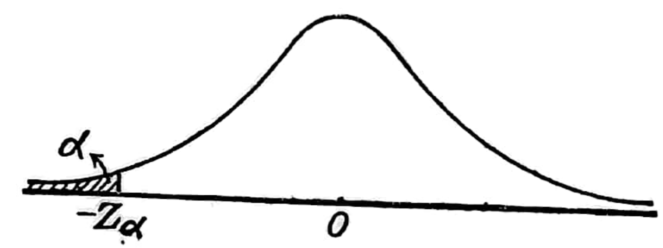
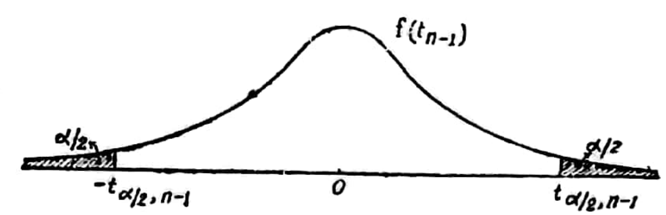
If r is the sample correlation coefficient of a random sample of size n from a bivariate normal population and if ρ is the population correlation coefficient then $\frac{1}{2}\log_e\frac{1+r}{1-r}$ is approximately normally distributed with mean $\mu=\frac{1}{2}\log_e\frac{1+\rho}{1-\rho}$ and with the standard deviation $\sigma=\frac{1}{\sqrt{n-3}}$. This property enables us to carry out the test in problem 13 of the following table.

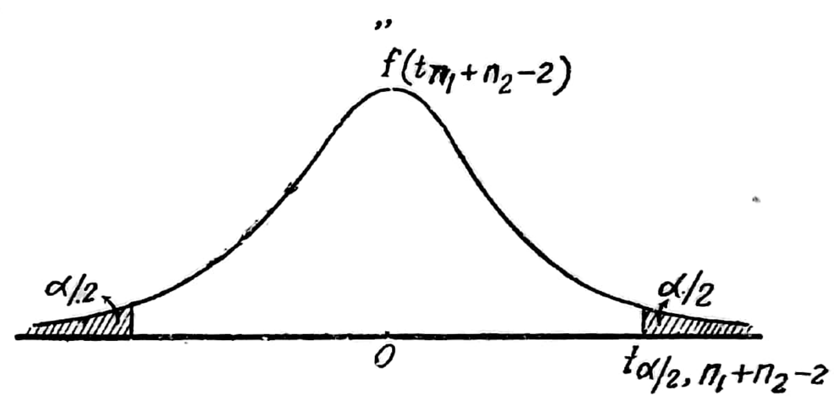
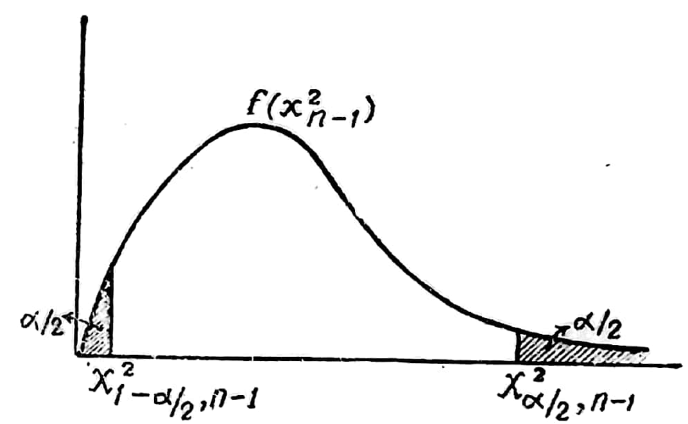
$$\sum_{x=0}^{K_{1-\alpha/2}} \binom{n}{x} p_0^x\, q_0^{n-x} \leqslant \alpha/2$$

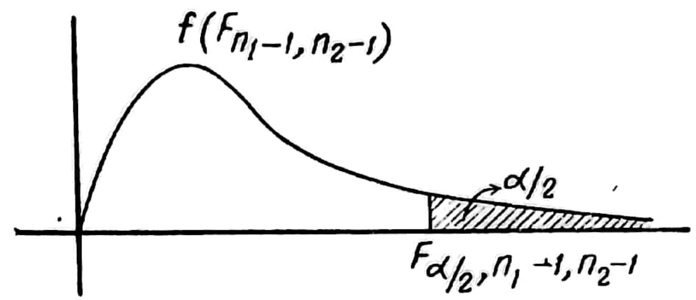
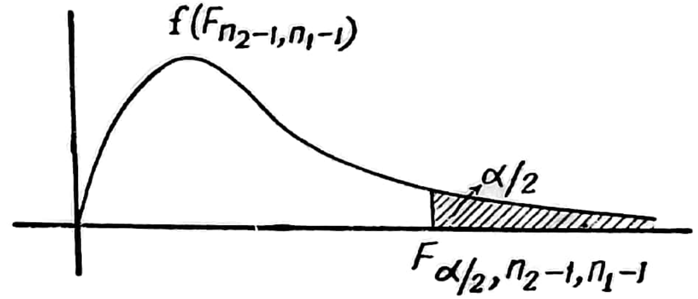
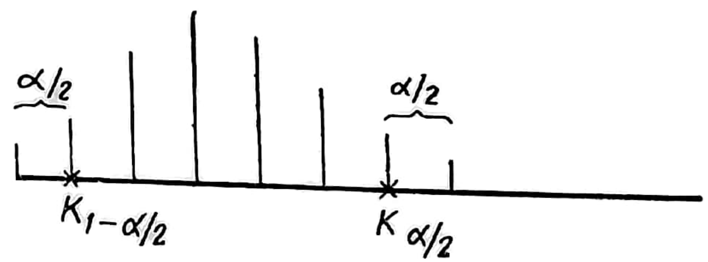
and

$$\sum_{x=K_{\alpha/2}}^{n} \binom{n}{x} p_0^x\, q_0^{n-x} \leqslant \alpha/2$$

where $q_0=1-p_0$ and $K_{1-\alpha/2}$ and $K_{\alpha/2}$ are the nearest integers which satisfy the above inequalities.

	Population	Hypothesis	Statistic	Criterion—reject H_i if	Illustration
1(*a*)	$N(\mu, \sigma)$ σ—known	$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} : N(0, 1)$	$\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \geqslant Z_\alpha$	
1(*b*)	„	$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	„	„ $\leqslant -Z_\alpha$	
1(*c*)	„	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	„	$\left\lvert\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right\rvert \geqslant Z_{\alpha/2}$	
2.	Any population $\mu < \infty, \sigma < \infty$ σ—known n—large	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	„ approx.	„	„
3.	$N(\mu, \sigma)$ σ—unknown $n \geqslant 30$	„	$\dfrac{\bar{X}-\mu}{S/\sqrt{n}} : N(0, 1)$ approx.	$\left\lvert\dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}\right\rvert \geqslant Z_{\alpha/2}$	„
4.	$N(\mu, \sigma)$ σ—unknown $n < 30$	„	$\dfrac{\bar{X}-\mu}{S'/\sqrt{n}} : t_{n-1}$	$\left\lvert\dfrac{\bar{x}-\mu_0}{s'/\sqrt{n}}\right\rvert \geqslant t_{\alpha/2,\, n-1}$	

	Population	Hypothesis	Statistic	Criterion—reject H_0 if	Illustration
5.	$N(\mu_1\ \sigma_1)$ and $N(\mu_2, \sigma_2)$ σ_1, σ_2—known	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$	$\dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} : N(0,1)$	$\left\lvert \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \right\rvert \geqslant Z_{\alpha/2}$	as in 1 (c)
6.	$N(\mu_1, \sigma_1)$ and $N(\mu_2\ \sigma_2)$ σ_1, σ_2 unknown $n_1 \geqslant 30, n_2 \geqslant 30$	"	$\dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ N(0, 1) approx.	$\left\lvert \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \right\rvert \geqslant Z_{\alpha/2}$	" 
7.	$N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$ $\sigma_1 = \sigma_2 = \sigma$ (unknown) $n_1 + n_2 - 2 \geqslant 30$	"	$\dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$: N(0, 1) approx.	$\left\lvert \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \right\rvert \geqslant Z_{\alpha/2}$	
8.	$N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$ σ—unknown $n_1 + n_2 - 2 < 30$	"	$\dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} : t_{n_1+n_2-2}$	$\left\lvert \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \right\rvert \geqslant t_{\alpha/2, n_1+n_2-2}$	
9.	$N(\mu, \sigma)$	$H_0 : \sigma = \sigma_0$ $H_1 : \sigma \neq \sigma_0$	$\dfrac{nS^2}{n-1} : \chi^2_{n-1}$	$\dfrac{ns^2}{\sigma_0^2} \leqslant \chi^2_{1-\alpha/2, n-1}$ or $\geqslant \chi^2_{\alpha/2, n-1}$	

10.	$N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$	H_0 ; $\sigma_1=\sigma_2$ $H_1 : \sigma_1 \neq \sigma_2$	$\dfrac{n_1 S_1^2 \Big/ \sigma_1^2\ (n_1-1)}{n_2 S_2^2 \Big/ \sigma_2^2\ (n_2-1)}$. $F_{n_1-1,\, n_2-1}$	$\dfrac{n_1 s_1^2 \Big/ (n_1-1)}{n_2 s_2^2 \Big/ (n_2-1)}$ $\geqslant F_{\alpha/2,\, n_1-1,\, n_2-1}$ if $s_1 \geqslant s_2$ $\dfrac{n s_{22}^2 \Big/ (n_2-1)}{n_1 s_1^4 \Big/ (n_1-1)}$ $\geqslant F_{\alpha/2,\, n_2-1,\, n_1-1}$ if $s_2 \geqslant s_1$	
11.	Binomial parameters n and p n—large $np>5$, $nq>5$	$H_0 : p=p_0$ $H_1 : p \neq p_0$	$\dfrac{X-np}{\sqrt{npq}}$: N(0, 1) approx.	$\dfrac{x-np_0}{\sqrt{np_0q_0}} \geqslant Z_{\alpha/2}$	as in 1 (c)
12.	Binomial parameters n and p n—small	,,	$X : \binom{n}{x} p^x q^{n-x}$	$x \leqslant K_{1-\alpha/2}$ or $\geqslant K_{\alpha/2}$	
23.	Bivariate normal with correlation ρ. r—sample correlation coefficient	$H_0 : \rho=0$ $H_1 : \rho \neq 0$	$\dfrac{\frac{1}{2}\log_e \frac{1+r}{1-r} - \frac{1}{2}\log \frac{1+\rho}{1-\rho}}{\left(\frac{1}{\sqrt{n-3}}\right)}$: N(0, 1) approx.	$\left\lvert \dfrac{\frac{1}{2}\log \frac{1+r}{1-r}}{\left(\frac{1}{\sqrt{n-3}}\right)} \right\rvert \geqslant Z_{\alpha/2}$	as in 1 (c)

BIBLIOGRAPHY

1. M.G. **Kendall** and A. **Stuart**. The Advanced Theory of Statistics. Vol. 2, Charles Griffin and Co., London, 1961.

2. J.F. **Kenney** and E.S. **Keeping**. Mathematics of Statistics, part 2, N.Y., D. Van Nonstrand Co. Inc. 1951.

3. E.L. **Lehmann**. Testing Statistical Hypotheses. N. Y. John Wiley and Sons, Inc. 1959.

4. H.M **Walker** and J. **Lev**. Statistical Inference. Holt, Rinehart and Winston, Inc., N.Y. 1953.

CHAPTER 11

CATEGORIZED DATA AND THE χ^2 STATISTIC

11.0. Introduction. In the last chapter we considered the problem of testing various hypotheses regarding the parameters of a distribution, based on the observations on an observable stochastic variable. Sometimes we will be interested in testing the association between two attributes or in testing whether a particular distribution is a good fit to a data etc. Problems of this nature are of practical importance. For example, suppose we have measured the heights of a random sample of university students. The data may be given as shown in the following table.

Height	55—60	60—65	65—70	70—75	75—80	80—85
Frequency	n_1	n_2	n_3	n_4	n_5	n_6

The data is classified into various classes and the number of individuals (frequencies) in the various classes are given. If we can find out the best fitting theoretical distribution to this frequency table we will be able to test various hypotheses regarding the distribution of heights of the university students under consideration. Such problems are called 'goodness of fit' problems.

Study of association between attributes or variables is of some use in many practical situations. One may be interested in seeing whether there is any association between the heights of persons and their intelligence, habit of wearing a tall hat and longevity of life, colour of eyes and the weight of persons, etc. If a data is given in the form of frequencies falling into different classes or categories then such a data is called a categorized data. These different categories may be characterized by measurable quantities such as temperature, height, length, etc., or non-measurable quantities like colour, state of existence etc. In the study of categorized data a χ^2 statistic is often very useful In this chapter we will consider the use of a χ^2 statistic in the analysis of categorized data. In the next chapter some other statistics will be considered.

11.1. GOODNESS OF FIT

The problem that is considered in this section is only a particular case of goodness of fit problems. We will not consider the existence of a best fitting distribution to a given data, but we will examine whether the data is compatible with a given theoretical distribution. In other words we will test the hypothesis that the data may be considered to be the observed values of (or the values assumed by) a stochastic variable having a specified distribution. Consider a multinomial probability situation (that is an experiment resulting in k mutually exclusive outcomes, with probabilities $p_1, p_2 \ldots p_k$

where $$\sum_{i=1}^{k} p_i=1.$$

Let $n_1, n_2 \ldots n_k$ be the observed outcomes in n independent trials (that is, $n=n_1+n_2+\ldots+n_k$). The joint probability function of $n_1, n_2 \ldots n_k$ is

$$f(n_1, n_2, \ldots n_k)=\frac{n!}{n_1!\, n_2! \ldots n_k!} p_1^{n1}\, p_2^{n2} \ldots p_k^{nk} \qquad (11.1)$$

$e_i=E(n_i)=np_i$ and the maximum likelihood estimates of p_i may be easily seen to be $\frac{n_i}{n}$ [obtained by maximizing $f(n_1, \ldots n_k)$ subject to the condition $\Sigma\, p_i=1$].

Let us examine whether a specified multinomial distribution is a good fit to the observed data. This is equivalent to the problem of testing the hypothesis

$$H_0 : p_i=p_{i0} \text{ for } i=1, 2, \ldots, k$$

where p_{i0} is the specified value of p_i. The likelihood ratio criterion λ in this case may be obtained as

$$-2 \log \lambda=2 \sum_{i=1}^{k} n_i \log \frac{n_i}{e_i} \qquad (11.2)$$

According to the Theorem 10.3, $-2 \log \lambda$ has a χ^2 distribution with degrees of freedom equal to the number of parameters specified by the null hypothesis H_0, when n is sufficiently large. Here the number of degrees of freedom is $k-1$ since $\Sigma\, p_i=1$ and only $k-1$ parameters are specified by H_0. $-2 \log \lambda$ in the equation (11.2) may be simplified to the form

$$\chi^2_{k-1}=\sum_{i=1}^{k} \frac{(n_i-e_i)^2}{e_i} \qquad (11.3)$$

The simplification involves some manipulations and the evaluation of some limits when $n \to \infty$, and is left to the reader. Therefore we may state the following theorems.

Theorem 11.1. If $n_1, n_2, \ldots, n_k$ and $e_1, e_2, \ldots e_k$ are the observed and theoretical frequencies in a multinomial probability situation respectively, then

$$\sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i} = \Sigma \frac{(\text{Observed frequency} - \text{Expected frequency})^2}{\text{Expected frequency}} \tag{11.4}$$

is approximately distributed as a χ^2 distribution with $k-1$ degrees of freedom when n (the number of trials) is sufficiently large.

(A good approximation is obtained when $k \geqslant 5$ and $e_i \geqslant 5$ for $i = 1, 2, \ldots k$).

Ex. 11.1.1. *The fishes caught by a man fishing at a certain spot in a lake are classified according to their weights in lbs and are given in the following table.*

Weight	*less than 1,*	*1—2*	*2—3*	*3—4*	*4—5*	*over 5*
frequency	*6*	*7*	*13*	*17*	*6*	*5*

Examine whether the data is compatible with the assumption that anyone fishing at this place will catch fishes in the ratio 1 : 1 : 2 : 3 : 1 : 1 in the various weight groups.

Sol. Total number of fishes caught $= n = 54$

The expected frequencies in the various weight groups are 6, 6, 12, 18, 6, 6.

Observed frequencies	$n_1=6$	$n_2=7$	$n_3=13$	$n_4=17$	$n_5=6$	$n_6=5$
Expected frequencies	$e_1=6$	$e_2=6$	$e_3=12$	$e_4=18$	$e_5=6$	$e_6=6$

$k =$ the number of classes $= 6$.

$$\chi^2_{k-1} = \chi^2_5 = \sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i}$$

$$= 0/6 + 1/6 + 1/12 + 1/18 + 0/6 + 1/6 = 0\ 47.$$

Comments. 0.47 is less than the tabled value of a χ^2 with 5 degrees of freedom at 5% level. The observed χ^2 is not in the critical region and hence we will accept the hypothesis that the data is compatible with the assumption that the ratios of the various weight groups are 1 : 1 : 2 : 3 : 1 : 1. In a problem if one class frequency is less than 5 then it may be combined with the

adjoining class so that the combined class frequency may be greater than 5. It may be noticed that the length of the class intervals, the units of measurements etc have nothing to do with the χ^2 statistic. The same χ^2 statistic may be used to test the hypothesis that $p_1=p_2=\dots=p_k$ where $p_1, p_2, \dots, p_k$ are the population proportions in k independent binomial populations.

If a theoretical distribution is fitted to a given data and if t parameters are estimated while fitting the distribution, then the χ^2 statistic of equation (11.4) may be used to test the goodness of fit, but the degrees of freedom will be $(k-1)-(t)$ where k is the number of classes.

Ex. 11.1.2. *The accident rates on a particular highway are given in the following table. Examine whether the data may be assumed to follow a Poisson distribution.*

Number of accidents	*0*	*1*	*2*	*3*	*4*	*5*	*6*
Number of days (frequencies)	*25*	*30*	*20*	*15*	*5*	*3*	*2*

Sol. We want to examine whether the distribution

$$f(x)=\frac{\lambda^x e^{-\lambda}}{x\,!}$$

is a good fit to the data, where x denotes the number of accidents. λ is to be estimated. The sample mean $\bar{x}$ may be used as an estimate for λ.

$$\bar{x}=[(0)(25)+(1)(30)+\dots+(6)(2)]/100=1.62$$

i.e. we want to examine the goodness of fit of

$$f(x)=\frac{(1.62)^x}{x\,!}\;e^{-1.62}.$$

The expected frequencies for various values of x are obtained by evaluating

$$100.\ \frac{(1.62)^x}{x\,!}\;e^{-1.62} \text{ for } x=0, 1, \dots 6.$$

Number of accidents (x)	*Frequencies* (n_i)	*Poisson probabilities* $f(x)$ *for* $\lambda=1.62$	*Expected frequencies* (e_i)
0	25	0.2019	20.19
1	30	0.3230	32.30
2	20	0.2584	25.84
3	15	0.1378	13.78
4	5	0.0551	5.51
5	3 } 10	0.0176	1.76 } 7.74
6	2	0.0047	.47

The expected frequencies for $x=6$ and for $x=5$ are less than 5. Further $e_6+e_5<5$. Hence they are combined with e_4. Therefore $k=5$ and the number of parameters estimated $=1$.

$$\chi^2_{k-l-t}=\chi^2_3=\sum_{i=1}^{5}\frac{(n_i-e_i)^2}{e_i}$$

$$\doteq(25-20.19)^2/(20.19)+(30-32.30)^2/(32.30)$$
$$+\ldots+(10-7.74)^2/(7.74)\doteq 3.09.$$

The tabulated value of a χ^2_3 at 5% level is greater than 3.09 and hence the observed χ^2_3 does not fall in the critical region. The hypothesis may be accepted. A Poisson distribution with $\lambda=1.62$ can be assumed to be a good fit.

Ex. 11.1.3. *The bust measurement of 80 women are given in the following table. The mean and the standard deviation of the measurements before the observations are classified, are 35 and 2 respectively. Test the goodness of fit of a normal distribution to this data.*

Measurements	*30 or less*	*31—32*	*33—34*	*35—36*	*37—38*	*39 or more*
Frequencies	*8*	*12*	*15*	*20*	*15*	*10*

Sol. The various classes may be assumed to be less than 30.5, 30.5 to 32.5, 32.5 to 34.5 etc. The expected frequencies in

the interval 30 or less = 80 times the probability that a normal variable with the parameters $\mu = 35$ and $\sigma = 2$ falls below 30.5.

$$= 80 \int_{-\infty}^{30.5} \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-35)^2}{2(4)}} dx = 80 \int_{-\infty}^{\frac{30.5-35}{2}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

$$= 80(0.0122) = 0{\cdot}976.$$

The total frequency in the class 32.5 or less

$$= 80 \int_{-\infty}^{\frac{32.5-35}{2}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = 8.448$$

$\therefore$ The frequency in the class 30.5 to 32.5

$= 8.448 - 0.976 = 7.472$ etc.

Class intervals	*Observed frequencies*	*Cumulated frequencies*	*Expected frequencies*
30.5 or less	8 } 20	0.976	0.976 } 8.448
30.5—32.5	12	8.448	7.472
32.5—34.5	15	32.104	23.656
34.5—36.5	20	61.872	29.768
36.5—38.5	15 } 25	76.792	14.920 } 18.128
38.5 or more	10		3.208

$k = 4$ and 2 parameters are estimated. Hence

$$\chi^2_{k-1-t} = \chi^2_1 = \Sigma \frac{(n_i - e_i)^2}{e_i} = (20 - 8.448)^2/(8.448)$$

$$+[(15 - 23.656)^2/(23.656) + (20 - 29.768)^2/(29.768)$$

$$+ (25 - 18.128)^2/(18.128) > 6.635$$

where 6.635 is the tabulated value of a χ^2_1 at 1% level. Hence the hypothesis is rejected.

Comments. Here $k = 4 < 5$. Hence the approximation of $\Sigma (n_i - e_i)^2/e_i$ to a χ^2_1 is not a good approximation. From these

examples it is seen that the applicability of a χ^2 test in testing goodness of fit is limited due to the need of classification, the restriction on k and e_i etc. Some exact tests will be considered in the next chapter.

Exercises

11.1. Test the goodness of fit of a multinomial distribution to the following data of the outcomes of an experiment of rolling a die 50 times. (The faces of the die are marked 1, 2,...6).

Face numbers	1	2	3	4	5	6
Frequencies	7	8	9	8	8	10

Assume that $p_1=p_2=\ldots=p_6=1/6$.

11.2. A historical monument is visited by 1000 people on a particular day. The exact categorization is given below.

Visitors from	North Amer.	South Amer.	Europe	Africa	Asia
Frequency	400	50	250	100	200

Is the data compatible with the assumption that the ratios will be 4 : 1 , 2 : 1 : 2 respectively.

11.3. The telephone calls received at an office switch-board are counted at every minute and the following data is obtained.

Number of calls (x)	0	1	2	3	4	5	6	7
Corresponding Number of one minute intervals	40	60	50	30	20	10	3	1

Test the goodness of fit of a Poisson distribution to the data.

11.4. The following is the classification of the families in a township according to the milk consumption.

Consumption of milk	5—10 units	11—15	16—22	23—27	28—32	33 or more
Number of families	200	180	170	140	100	3)

Test the goodness of fit of an exponential distribution to the data. The average consumption=18 units (calculated before classification).

11.5. The marks obtained by 100 students have a mean 70 with a standard deviation 5. The marks are classified and then given in the following table.

Marks	50 or less	51—60	61—70	71—80	81—90	91—100
Number of Students	5	10	30	25	20	10

Is the data compatible with the assumption that the marks are normally distributed. Can you generalize your findings ? If not, why ?

11.2. CONTINGENCY TABLES

The following is the data of waist measurements of 100 women classified according to their intelligence.

Measurements/ Intelligence	*Very intelligent*	*Average*	*Below average*
16 or less	10	8	8
17—18	6	7	8
19 - 20	7	8	7
21—22	5	6	5
23 or more	5	5	5

The numbers in the various cells are the frequencies, *i.e.* the number of measurements falling under the corresponding characteristics of classification. If the data is classified according to two or more characteristics (measureable or not) and is given in a frequency table then such a table is often called a contingency table. The study of association between two characteristics of classification is of some practical interest. For example we would like to test whether there is any association between heights and intelligence, oratorical talents and the quantity of food consumed, aptitude for Mathematics and the interest in games, etc. A χ^2 statistic can be conveniently used to test the independence of the classification in a two-way contingency table. Let $A_1, A_2, \ldots A_r$ and $B_1, B_2, \ldots B_s$ be the categories of the two characteristics under consideration and let the data be given as shown in the following table.

	B_1	B_2	...	B_s	*Total*
A_1	n_{11}	n_{12}	....	n_{1s}	$n_{1\cdot}$
A_2	n_{21}	n_{22}	....	n_{2s}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_r	n_{r1}	n_{r2}	....	n_{rs}	$n_{r\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$		$n_{\cdot s}$	$n_{\cdot\cdot}$

n_{ij} is the number of observations (frequency) corresponding to A_i and B_j.

$i=1, 2, \ldots, r$; $j=1, 2, \ldots s$; $\sum_{i=1}^{r} n_{ij}=n_{\cdot j}$;

$$\sum_{j=1}^{s} n_{ij}=n_{i\cdot} ; \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij}=n_{\cdot\cdot}$$

In the above notation, the summation with respect to a suffix is denoted by a dot.

Let p_{ij} be the probability of getting an observation in the $(ij)^{th}$ cell (i^{th} row, j^{th} column cell), then

$\sum_{j=1}^{s} p_{ij}=p_{i\cdot}=$the probability of getting an observation in the i^{th} row.

$\sum_{i=1}^{r} p_{ij}=p_{\cdot j}$ the probability of getting an observation in the j^{th} column.

If the classification is independent of one another then

$$p_{ij}=p_{i\cdot}p_{\cdot j}.\ i=1, 2, \ldots r \text{ and } j=1, 2, \ldots s.$$

Consider the hypothesis,

$H_0 : p_{ij}=p_{i\cdot}\, p_{\cdot j}$

(There is independence in the classification.)

$H_1 : p_{ij}\neq p_{i\cdot}\, p_{\cdot j}$ at least for one i and j

(There is no independence.)

Under H_0 expected frequency in the (ij^{th}) cell is

$$e_{ij}=(p_{i\cdot}\, p_{\cdot j})n_{\cdot\cdot}$$

Since $p_{i\cdot}$ and $p_{\cdot j}$ are unknown they may be estimated by

$$\hat{p}_{i\cdot}=\frac{n_{i\cdot}}{n_{\cdot\cdot}} \text{ and } \hat{p}_{\cdot j}=\frac{n_{\cdot j}}{n_{\cdot\cdot}}$$

Hence $$\hat{e}_{ij}=\frac{n_{i\cdot}}{n_{\cdot\cdot}}\cdot\frac{n_{\cdot j}}{n_{\cdot\cdot}}=\frac{n_{i\cdot}\, n_{\cdot j}}{n_{\cdot\cdot}}$$

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(n_{ij}-\hat{e}_{ij})^2}{\hat{e}_{ij}}=\chi^2_{(r-1)(s-1)} \text{ approximately.}$$

The degrees of freedom = total number of classes minus one minus the number of independent parameters estimated

$$=rs-1-(r+s-2)=(r-1)(s-1).$$

Since $\sum_{i=1}^{r} p_{i\cdot}=p_{\cdot\cdot}=1=\sum_{j=1}^{s} p_{\cdot j}$, only $r+s-2$ independent

parameters are estimated. The approximation is good if all the $e_{ij} \geqslant 5$ and $rs \geqslant 5$.

Ex. 11.2.1. *Check whether the data given in section 11.2 are compatible with the assumption that there is no association between intelligence and waist measurements in women.*

Sol. The data is reproduced in the following table.

	B_1	B_2	B_3	Total
A_1	$^{10}8.58$	$^{8}8.84$	$^{8}8.58$	26
A_2	$^{6}6.93$	$^{7}7.14$	$^{8}6.93$	21
A_3	$^{7}7.26$	$^{8}7.48$	$^{7}7.26$	22
A_4	$^{5}5.28$	$^{6}5.44$	$^{5}5.28$	16
A_5	$^{5}4.95$	$^{5}5.10$	$^{5}4.95$	15
Total	33	34	33	100

$\hat{e}_{11} = (26)(33)/100 = 8.58$

$\hat{e}_{12} = (26)(34)/100 = 8.84$ etc.

The estimated expected frequencies are given at the corner of each cell

$$(r-1)(s-1) = (4)(2) = 8.$$

$$\therefore \quad \chi_8^2 = \sum_{i=1}^{5} \sum_{j=1}^{3} \frac{(n_{ij} - \hat{e}_{ij})^2}{\hat{e}_{ij}} = (10-8.58)^2/(8.58)$$

$$+(6-6.93)^2/(6.93) + \ldots + (5-4.95)^2/(4.95) < 15.507$$

where 15.507 is the tabled value of a χ^2 with 8 degrees of freedom at 5% level. Hence we will accept the null hypothesis that there is no evidence of any association between the waist measurements and intelligence, based on the data given.

Comments. Even if our null hypothesis of independence is rejected we cannot generalize our results. We can only say that the data is not compatible with the hypothesis.

Analogous to the correlation coefficient between two variables some measures of association between the characteristics of categorization in a two-way contingency table, are suggested. Some of the commonly used measures are :

(1) Square contingency

$$X^2 = n.. \left[\sum_{i=1}^{r} \sum_{j=1}^{s} \frac{n^2_{ij}}{n_{i.}\, n_{.j}} - 1 \right]$$

(2) K. Pearson's coefficient of contingency P

$$P = \left(\frac{\chi^2}{n.. + \chi^2} \right)^{1/2}$$

where χ^2 is the χ^2 calculated from a contingency table under the assumption of independence and $n..$ is the total frequency. Evidently $P=0 \Leftrightarrow \chi^2=0$ and $0 \leqslant P \leqslant 1$. When there is complete independence $P=0$ and *vice versa.*

Exercises

11.6. The following table gives the categorized data classifying 100 people according to their intelligence and their mood upon getting up on a particular morning. Is there any evidence of association between these characteristics from the data given ?

Mood Intelligence	*Highly intelligent*	*Intelligent*	*Average*	*Below Average*
Good	15	12	10	5
Tolerable	10	10	5	5
Intolerable	5	10	8	5

Calculate a coefficient of contingency. Can you generalize your inference to all the people in a country ?

11.7. The following are the data obtained from an experiment conducted to study the association between the power of concentration (measured in terms of time units) and the ability in Statistics

Power of Concentration

		1 hour or less	1—2	2—3	3—4	4 or more
Ability in Statistics	High	5	6	7	9	10
	Average	5	6	6	6	8
	Low	5	8	9	5	5

Test for independence in this classification at, 5% level.

11.8. The following is a classification of 95 people according to their heights and weights. Test for independence in the classification. Also calculate the Pearson's coefficient of contingency.

Weights / Heights	55—60	61—65	66 or more
120 or less	10	8	5
121—135	8	9	10
136 or more	5	6	14

BIBLIOGRAPHY

1. H. **Cramer.** Mathematical Methods of Statistics Princeton University Press, N.J. 1946.

2. M.G. **Kendall** and A. **Stuart.** The Advanced Theory of Statistics, Vol. 2. Charles Griffin and Co. Ltd., London, 1961.

3. G.U. **Yule** and M.G. **Kendall.** An Introduction to the Theory of Statistics. 14th ed., N.Y. Hafner Publishing Co. 1950.

ANALYSIS OF DISPERSION

12.0. Introduction. In the last few chapters we considered some of the problems of statistical inference, namely, estimation and testing of statistical hypotheses. In this chapter we will present a unified theory for estimation and testing procedures. The concept of a statistical population, either defined by a set of given data or by a stochastic variable, is already familiar to the reader. We will define a measure of dispersion in a univariate population.

12.1. A MEASURE OF DISPERSION IN A GIVEN DATA

Let $x_1, x_2, \ldots x_n$ be the elements of the given population. A measure of scatter of the elements from any point of reference (a point of location) may be defined by the following axioms or desirable properties. Let m be a point of reference and let

$$d_i = x_i - m \text{ for } i = 1, 2, \ldots n$$

Any function D of $d_1, d_2, \ldots d_n$, satisfying the following conditions can be taken as a measure of dispersion in $x_1, x_2, \ldots, x_n$ from m.

a_1 $\quad D(d_1, d_2, \ldots d_n) > 0$ and $D = 0 \Leftrightarrow d_i = 0$ for $i = 1, 2, \ldots n$

a_2 $\quad D(ad_1, ad_2, \ldots ad_n) = | a | D(d_1, d_2, \ldots d_n)$ where a is a scalar quantity and $| a |$ is the absolute value of a.

a_2 $\quad D(d_1 + f_1, d_2 + f_2, \ldots, d_n + f_n) \leqslant D(d_1, d_2, \ldots d_n) + D(f_1, f_2, \ldots f_n)$ where $(f_1, f_2, \ldots f_n)$ is similarly constructed from another population $(y_1, y_2, \ldots y_n)$.

a_4. $\quad D(b_1, \ldots, b_n) = 1$ when $| b_i | = 1$ for $i = 1, 2, \ldots, n$.

Axiom a_1 suggests that the measure is always a positive quantity and is equal to zero if and only if all the $x_1, \ldots, x_n$ coincide with m. a_2 says that if the elements are scaled by a scalar quantity a, then the measure itself is scaled by the magnitude of a. Axiom a_3 is equivalent to the statement that the dispersion of a sum is less than or equal to the sum of the dispersions. If all the elements are one unit away from m we would like to have a measure of scatter also equal to unity or if the elements are c units away from m we would like to have the measure

to be c. Axiom a_4 together with a_2 gives this result. The following are some of the examples for such a function.

$$D_1=\left\{ (1/n)\ \Sigma_{i=1}^{n}\ |d_i|^r \right]^{1/r} \text{ for } r\geqslant 1. \tag{12.1}$$

$$D_2=\max_i\ |\ d_i\ | \tag{12.2}$$

$$D_3=\left[\ c_1\ d_1^2+\dots+c_n\ d_n^2\ \right]^{1/2} \text{ where } c_i>0 \text{ for } i=1, 2,\dots, n\ ;$$

$$c_1+\dots+c_n=1. \tag{12.3}$$

$$D_1 \text{ for } r=1 \text{ equal to } (1/n)\ \Sigma_{i=1}^{n}\ |\ x_i-m\ |\ . \tag{12.4}$$

This is the usual measure of mean absolute deviation from m.

$$D_1 \text{ for } r=2 \text{ is equal to } \left[\ (1/n)\ \Sigma_{i=1}^{n}\ |\ x_i-m\ |^2\ \right]^{1/2}. \tag{12.5}$$

This is the usual measure of standard deviation when m is the arithmetic mean of the x's. The constants $c_1, c_2,\dots c_n$ in D_3 satisfy the conditions for a probability measure and hence we can extend the concept of dispersion to a population defined by a stochastic variable.

Let a population be defined by a stochastic variable X' ; let m be a point of location so that

$$X=X'-m.$$

A measure of dispersion D in this population, from the point m, may be defined by the following axioms :—

a_1' . $D(X)>0$ and $D=0 \Leftrightarrow X=0$ almost surely.

a_2' $D(aX)=\ |\ a\ |\ D(X)$ where a is a scalar quantity.

a_3' . $D(X+Y)\leqslant D(X)+D(Y)$ where $Y=Y'-m$ and Y' represents another population.

a_4' . $D(X)=1$ if $|\ x\ |\ =1$ almost surely.

The following are some of the functions D which satisfy the conditions a_1' , a_2' , a_3' and a_4' :

$$D_4 = \left[E_X |X|^r \right]^{1/r} \quad \text{for } r \geqslant 1 \tag{12.7}$$

$$D_5 = \sup_X |X| \tag{12.6}$$

where E_X denotes mathematical expectation with respect to X and sup means the supremum or the maximum of $|X|$.

A number of other measures may be constructed in a similar way.

12.2. THE PRINCIPLE OF MINIMUM DISPERSION

This is a general principle which enables us to get some criteria for estimation of parameters and testing of statistical hypotheses. Suppose that a parameter θ is estimated by an estimator $\hat{\theta}$, then $\hat{\theta}-\theta$ may be taken as an error in the estimation. If there exists an estimator which minimizes the dispersion $D(\hat{\theta}-\theta)$ then such an estimator may be called the best, in the sense of minimum dispersion. In general if θ designates a correct situation and if $\hat{\theta}$ designates a statistic designed for θ and if θ is evaluated by minimizing any measure of dispersion $D(\hat{\theta}-\theta)$ the principle is called the principle of minimum dispersion. This principle will be illustrated in the following sections.

12.3. THE PRINCIPLE OF LEAST SQUARES

Let $y_1, y_2, \ldots y_n$ be a given data and let $y_{1\theta}, y_{2\theta}, \ldots, y_{n\theta}$ be the hypothetical values of $y_1, \ldots y_n$ respectively, then $e_i = y_i - y_{i\theta}$ may be called the error in the i^{th} observation, for $i=1, 2, \ldots n$. If we take a measure of dispersion as D_1 for $r=2$, that is,

$$D = \left[\frac{1}{n} \sum_{i=1}^{n} e_i^2 \right]^{1/2} = \left[\frac{1}{n} \sum_{i=1}^{n} (y_i - y_{i\theta})^2 \right]^{1/2} \tag{12.8}$$

and if the unknown values denoted by θ, are estimated by minimizing D, then the estimation procedure is called the principle of least squares. For example let $x_1, x_2, \ldots, x_n$ denote the height measurements of n persons and let $y_1, y_2, \ldots, y_n$ be their weights. If we assume that there is a linear relationship between the height and weight such that if the height is given the weight can be estimated by using a relationship of the form $y=a+bx$ where y and x denote weight and height respectively. Here our hypothetical model under the assumptions of linear relationship between x and y is $a+bx_i$ for y_i. The error is y_i-a-bx_i where a and b are unknowns. We can estimate a and b by the principle of least squares

$$D = \left[\frac{1}{n} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \right]^{1/2} \tag{12.9}$$

Minimization of D with respect to a and b is equivalent to minimization of

$$L = \sum_{i-1}^{n} (y_i - a - bx_i)^2 \qquad (12.10)$$

By using the principles of calcu lus

$$\frac{\partial L}{\partial a} = 0 \text{ and } \frac{\partial L}{\partial b} = 0$$

corresponding to the maximum or minimum values. These equations are called the normal equations.

i.e., $$\frac{\partial L}{\partial a} = 0 \Rightarrow -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - a - bx_i) = 0 \qquad (12.11)$$

$$\frac{\partial L}{\partial b} = 0, \Rightarrow -2 \sum_{i=1}^{n} x_i(y_i - a - bx_i) = 0, \Rightarrow \sum_{i=1}^{n} x_i(y_i - a - bx_i) = 0$$

Equations (12.11) and (12 12) yield,

$$\sum_{i=1}^{n} y_i - na - b \sum_{i=1}^{n} x_i = 0 \qquad (12.13)$$

$$\sum_{i=1}^{n} x_i y_i - a \sum_{i=1}^{n} x_i - b \sum_{i=1}^{n} x_i^2 = 0 \qquad (12.14)$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} - \text{ and } \hat{b} = \frac{\Sigma x_i y_i / n - \bar{x}\,\bar{y}}{\Sigma x_i^2 / n - \bar{x}^2} = \frac{\text{Cov }(x, y)}{\text{Var }(x)} \qquad (12.15)$$

where $\hat{a}$ and $\hat{b}$ denote the estimated values and Cov (x, y) and Var (x) denote the observed sample covariance and the variance of X, respectively. It may be easily seen that $\hat{a}$ and $\hat{b}$ minimize L.

Similarly if we have a set of observed values $y_1, y_2, \ldots y_n$ and hypothetical values $\phi_1(\theta), \phi_2(\theta), \ldots, \phi_n(\theta)$ respectively and if the parameters in $\phi(\theta)$ are estimated by minimizing

$$L = \sum_{i=1}^{n} [y_i - \phi_i(\theta)]^2 \qquad (12.16)$$

with respect to the parameters. the principle is called the principle of least squares and the corresponding estimates are called the least square estimates.

Ex. 12.3.1. *Assuming a linear relationship of the form $y=a+bx$ between heights and weights, fit a straight line to the following data.*

Height (x)	64	65	63	66	67
Weight (y)	125	130	120	140	150

Sol. If $a+b\,x_i$ is the value set up for y_i the error

$$e_i=y_i-a-b\,x_i$$

Minimizing the dispersion in e_i for $i=1, 2, 3, 4, 5$ by minimizing the dispersion D_1 for $r=2$ or by using the principle of least squares,

$$L=\sum_{i=1}^{5}(y_i-a-b\,x_i)^2 \tag{12.17}$$

and

$$\frac{\partial L}{\partial a}=0=\frac{\partial L}{\partial b}$$

$$\Rightarrow \hat{a}=\bar{y}-\hat{b}\,\bar{x} \text{ and } \hat{b}=\frac{\Sigma\, x_i\, y_i/n-\bar{x}\bar{y}}{\Sigma\, x_i^2/n-\bar{x}^2} \quad \left[\text{by the equation } (12.15)\right]$$

But $\bar{x}=(64+65+63+55+67)/5=65$; $\bar{y}=133$

$$\Sigma\, x_i\, y_i/n-\bar{x}\bar{y}=15$$

$$\Sigma\, x_i^2\Big/n-\bar{x}^2=2,$$

$$\hat{b}=7.5 \qquad \hat{a}=-354.5$$

$\therefore$ The estimated equation is

$$y=-354.5+7.5x.$$

Comments. If more observations were available the estimates $\hat{a}$ and $\hat{b}$ would be different. If there is an exact linear relationship $y=a+bx$, the two parameters may be evaluated by using two pairs of values for x and y and if we substitute any other pair of values for x and y the equation should be satisfied. Under our assumption of a linear relationship we can only estimate the parameters. Our assumption may be tested by using a 'goodness of fit' test.

Ex. 12.32. *Fit a curve $y=ab^x$ to the data in Ex. 12.3.1.*

Sol. $y=ab^x, \Rightarrow \log y=\log a+x\log b$ (12.18)

i.e., $Y = A + Bx$ where $Y = \log y$, $B = \log b$ and $A = \log a$.

By using the results in the equation (12.15) (12.19)

$$\hat{A} = \overline{Y} - \hat{B}\bar{x} \text{ and } \hat{B} = \frac{\Sigma x_i Y_i / n - \bar{x}\overline{Y}}{\Sigma x_i^2 / n - \bar{x}^2} \tag{12.20}$$

$\hat{A} = 0.5429$ $\hat{B} = 0.0243$

$\therefore$ $\hat{a} = 1.3490$ $\hat{b} = 1.106$

$\therefore$ The estimated equation is $y = (1.3490)(1.106)^x$

Comments. From these examples it is clear that any curve may be fitted to a given data. The goodness of fit depends on the assumption we make. Then our assumption, that the relationship is of a particular nature, can be tested. Some more test statistics for 'goodness of fit' will be introduced later. If our assumption is acceptable, we can predict y based on any observed x ; *i.e.*, $y_0 = a + b^{x_0}$ where x_0 is the observed x, and y_0 is the predicted value of y. The error in this prediction depends on the validity of our model, the error of measurement in x_0 etc.

12.31. The scatter diagram. If we have a data of paired observations on two variables x and y (for example, height and weight measurements of n persons) we can plot the points (x, y) in a two dimensional space. Such a diagram is called a scatter diagram. An illustration is given in Fig. 12.1.

x	$x_1, x_2, \ldots\ldots\ldots x_n$
y	$y_1, y_2, \ldots\ldots\ldots y_n$

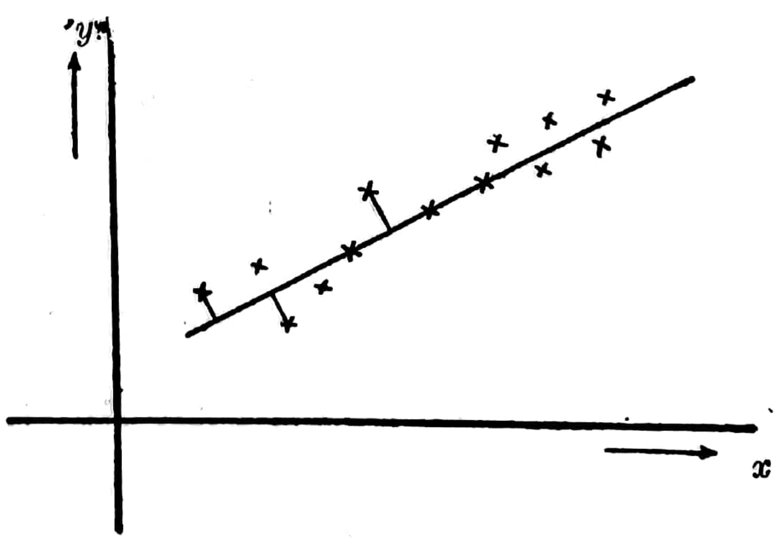


Fig. 12.1.

From such a diagrammatic representation it is easy to obtain some idea about the best fitting model to the data. In Fig. 12.1. a straight line seems to be a good fit. By the principle of least squares we minimize the sum of the squared distances of the points from the curve and estimate the unknown quantities in the model. The goodness of fit may also be tested by assuming some known distribution for the error in the model. If there are three variables the scatter diagram lies in a three dimensional space and

in general if there are observations on k-variables the scatter diagram will be in a k-dimensional space. An application of the principle of minimum dispersion, especially the principle of least squares, will be considered in the next section.

Exercises

12.1. A statistical population is given by the following data, 20, 25, 22, 27, 30, 35, 32, 31, 30, 26, 25, 27, 29, 24, 23, 22, 20, 21, 23, 26. Obtain the following measures of dispersion from the point 25. (1) D_1 for $r=1$, (2) D_1 for $r=2$. (3) D_2 (see section 12.1).

12.2. If a statistical population is defined by the stochastic variable X with the density function,

$$f(x)=\begin{cases} e^{-x} & \text{for } x>0 \\ 0 & \text{elsewhere,} \end{cases}$$

obtain the following measures of dispersion from the point zero ;

(1) D_4 for $r=1$, (2) D_4 for $r=2$, (3) D_5 (see section 12.1).

12.3. Obtain the least square normal equations in each case if the following curves are fitted to the data. (1) $y=a+b(x-\bar{x})$, (2) $y=a+bx+cx^2$, (3) $y=ab^x$, (4) $y=ax+b/x$, (5) $xy^a=b$, (6) $x=aye^{-by}$, (7) $x=e^{-ay-by^2-c}$, where a, b, c are constants and the data is given as follows :

y	y_1	y_2	........,	y_n
x	x_1	x_2	........,	x_n

12.4. Fit the curves, (1) $y=a+bx+cx^2$, (2) $y=ax+b/x$, (3) $x=aye^{-by}$, to the following data.

y	10	12	13	15	17	18	19	21	22	24
x	8	9	10	11	13	15	16	18	19	20

12.5. Draw the scatter diagram for the data in problem 12.4.

12.4. LINEAR REGRESSION

This is a simple problem where we can apply the principle of least squares to estimate linear relationships between stochastic variables. If X and Y are two stochastic variables, the conditional expectation of Y given X, that is $E(Y \mid X)$ is called the regression of Y on X. Similarly $E(X \mid Y)$ is called the regression of X on Y. These regressions need not be linear. In section 5.32.4 we had seen that if X and Y have a joint normal probability distribution then the regressions $E(Y \mid X)$ and $E(X \mid Y)$ are linear. If $E(Y \mid X)$ is linear then Y is said to be a linear regression on X and *vice versa*. If $X_1, \ldots, X_n$ are n stochastic variables then $E(X_1 \mid X_2, \ldots, X_n)$ is called the regression of X_1 on $X_2, \ldots, X_n$. If this multivariate linear regression is linear then it is called a multivariate linear regression.

Multivariate regression of X_i on the other variables, for $i=1$, 2, ..., n may be similarly defined. In this section we will consider linear regressions, that is, the cases when $E(X_1 \mid X_2, \ldots, X_n)$ etc., are linear. For example a linear regression of Y on X may be written as,

$$E(Y \mid X)=a+bx \tag{12.21}$$

where a and b are constants which are also called the regression coefficients. In general a linear regression of X_1, on X_2. X_3, ..., X_n may be written as

$$E(X_1 \mid X_2, X_3, \ldots, X_n)=a_1+a_2\, x_2+a_3\, x_3+\ldots+a_n x_n \tag{12.22}$$

where a_1, a_2, ..., a_n are constant regression coefficients. 'Conditional expectation' is discussed in section 5.23.

Theorem 12.1. If X and Y are two stochastic variables then $E(Y)=E_X[E_Y(Y \mid X)]$, where E_X and E_Y denote expectation with respect to X and Y respectively and the conditional expectation $E_Y(Y \mid X)$ is treated as a function of X.

Proof. Let X and Y be continuous and let $f(x, y)$, $f(x)$, $g(y)$ be the joint density and marginal densities respectively.

$$E_Y(Y \mid X)=\int_y y.\, h(y \mid x)dx=\int_y y\frac{f(x, y)}{f(x)}\, dy.$$

where $h(y|x)$ denotes the conditional distribution of Y given X.

$$=\frac{1}{f(x)}\int_y y\, f(x, y)dy \tag{12.23}$$

$$E_X[E_Y(Y \mid X)]=\int_x\left[\frac{1}{f(x)}\int_y y\, f(x, y)dy\right]f(x)dx$$

$$=\int_x\int_y y\, f(x, y)dy\, dx$$

$$=\int_y y\left[\int_x f(x, y)dx\right]dy$$

$$=\int_y y\, g(y)dy=E(Y). \tag{12.24}$$

The proof when X and Y are discrete or when one variable is discrete and the other is continuous, is left to the reader.

Theorem 12.2. If the regression of Y on X is linear, that is, if $E(Y \mid X)=a+bx$, then

$$a=\mu_2-\frac{\sigma_{12}}{\sigma_1^2}\mu_1 \quad \text{and} \quad b=\frac{\sigma_{12}}{\sigma_1^2}$$

where μ_1, σ_1^2 and μ_2, σ_2^2 are the means and variances of X and Y respectively and σ_{12} is the covariance between X and Y.

Proof. $E(Y)=E_X[E_Y(Y \mid X)]$ (Theorem 12.1)

$$=E_X[a+bX].$$

i.e.,
$$\mu_2=a+b\,\mu_1. \tag{12.25}$$

$$E(XY)=E_X[E_Y(XY \mid X)]=E_X[XE_Y(Y \mid X)]$$
$$=E[X(a+bX)]=aE(X)+b.E(X^2)$$

i.e.,
$$E(XY)=a\,.\,\mu_1+b\,.\,E(X^2) \tag{12.26}$$

Solving (12.25) and (12.26),

$$b=\frac{\sigma_{12}}{\sigma_1^2}$$

$$\left(\text{since } \sigma_{12}=E(XY)-\mu_1\mu_2 \text{ and } \sigma_1^2=E(X^2)-\mu_1^2\right)$$

$$a=\mu_2-\frac{\sigma_{12}}{\sigma_1^2}\mu_1$$

Hence
$$E(Y \mid X)=\mu_2+\frac{\sigma_{12}}{\sigma_1^2}(y-\mu_1) \tag{12.27}$$

Similarly
$$E(X \mid Y)=\mu_1+\frac{\sigma_{12}}{\sigma_2^2}(y-\mu_2) \tag{12.28}$$

12.41. Least Square Estimation of the Linear Regression Equations. Assuming that the regression of Y on X is linear, we can estimate the regression coefficients by setting up a model $a+bx$ for y or by letting

$$y=a+bx+e \tag{12.29}$$

where e is the error in the model. a and b may be estimated by the principle of least squares. These estimates are given in (12.15).

$$\hat{a}=\bar{y}-\hat{b}\bar{x} \text{ and } \hat{b}=\frac{\text{Cov }(x, y)}{\text{Var }(x)}.$$

$\therefore$ The estimated equation is

$$y-\bar{y}=\frac{\text{Cov }(x, y)}{\text{Var }(x)}(x-\bar{x}). \tag{12.30}$$

If there is linear regression of X on Y or if

$$E(X \mid Y)=c+dy$$

where c and d are constants, we can estimate the constants by setting up a linear model for x in the form

$$x=c+dy+e' \tag{12.31}$$

where e' is the error in the model. By applying the principle of least squares, c and d may be estimated. These estimates are easily seen to be

$$\hat{c}=\bar{x}-\frac{\text{Cov }(x, y)}{\text{Var }(y)}\cdot\bar{y}$$

and
$$\hat{d}=\frac{\text{Cov }(x, y)}{\text{Var }(y)} \tag{12.32}$$

The estimated equation is,

$$x-\bar{x}=\frac{\text{Cov }(x, y)}{\text{Var }(y)}(y-\bar{y}). \tag{12.33}$$

If there are more than two variables and if there is linear or non-linear regression the principle of least squares may be used to estimate the regression coefficients. If we assume distribution for e in the equation (12.29) we can construct confidence intervals for the regression coefficients.

For example, if e_i is assumed to have a normal distribution $N(0, \sigma)$ for $i=1, 2, \ldots, n$ then the likelihood function

$$L=f(e_1, e_2, \ldots e_n)=\prod_{i=1}^{n} f(e_i)=\frac{1}{\sigma^n(\sqrt{2\pi})^n} e^{-\sum_{i=1}^{n}(y_i-a-bx_i)^2/2\sigma^2}$$

a and b can be estimated by using the principle of maximum likelihood. Confidence intervals for a and b can be obtained from the distributions of $\hat{a}$ and $\hat{b}$. The assumption of $e_i : N(0, \sigma)$ in the equation (12.29) is equivalent to the assumption that

$$Y_i : N(a+bx_i, \sigma) \text{ and } x_i \text{ are constants.}$$

Ex. 12.4.1. *Obtain the regression of X on Y and Z, where the joint density function of X, Y and Z is given as,*

$$f(x, y, z)=\tfrac{1}{3}(x+y)e^{-z} \text{ for } 0<x<1,\ 0<y<2,\ z>0$$
$$=0 \text{ elsewhere}$$

Sol. $$f(y, z)=\int_0^1 f(x, y, z)\, dx$$

$$=\tfrac{1}{3}\int_0^1 (x+y)\, e^{-z}\, dx=\tfrac{1}{3}\,(y+1/2)\, e^{-z}. \qquad (12.34)$$

$$f(x \mid y, z)=f(x, y, z)/f(y, z)=\tfrac{1}{3}\ \frac{(x+y)}{(y+1/2)}. \qquad (12.35)$$

The regression of X on Y and Z is E(X | Y, Z)

$$=\int_0^1 x.f\,(x \mid y, z)\, dx=\tfrac{1}{3}\int_0^1 x.\ \frac{(x+y)}{(y+1/2)}\, dx$$

$$=\frac{1}{9}\cdot\frac{2+3y}{1+2y}. \qquad (12.36)$$

Therefore the regression curve is

$$E(X \mid Y, Z)=\frac{1}{9}\cdot\frac{(2+3y)}{(1+2y)} \qquad (12.37)$$

12.5. MINIMAX PROCEDURES

The following is a note on minimax procedures. We have seen that if the parameters are estimated by minimizing D_1 for $r=2$ (see section 12 1), the corresponding procedure is called the least square method. If $x_1,\ldots,x_n$ are the observed values with $y_{1\theta},\ldots, y_{n\theta}$ being the corresponding hypothetical values and if the parameters are estimated by minimizing the maximum error ($\min_\theta D_2$ or $\min_\theta \max_i \mid x_i-y_{i\theta} \mid$ see section 12.1), the estimation procedure is called the minimax procedure If $\hat{\theta}$ is selected as an estimator for θ then

$$E_{\hat{\theta}} \mid \hat{\theta}-\theta \mid,\ \{ E_{\hat{\theta}} \mid \hat{\theta}-\theta \mid^{p}\}^{1/p} \text{ for } p\geqslant 1, \text{ etc.}$$

are some of the measures of dispersion in $\hat{\theta}-\theta$, where $E_{\hat{\theta}}$ denotes the expectation with respect to the stochastic variable $\hat{\theta}$. The maximum dispersion can be denoted by $\max_\theta D(\hat{\theta}-\theta)$ or $\sup_\theta D(\hat{\theta}-\theta)$ where $D(\hat{\theta}-\theta)$ is any measure of dispersion in $(\hat{\theta}-\theta)$. For

different estimators $\sup_{\theta} D(\hat{\theta}-\theta)$ may be different. If an estimator $\hat{\theta}$ is such that it minimizes the maximum dispersion, that is, sup $D(\hat{\theta}-\theta)$ is a minimum, then $\hat{\theta}$ is called the minimax estimator, and the corresponding procedure of selecting an estimator $\hat{\theta}$ is also called the minimax procedure. Minimax estimators need not always exist.

Ex. 12.5.1. *1, 3 and 6 is a random sample of size 3 from a population* $f(x, \theta)=1/\theta$ *for* $0<x<\theta$ *and is zero elsewhere. Obtain a minimax estimate of* θ.

Sol. Under this model the expected or the hypothetical value of any observation is

$$E(X)=\int_0^{\theta} x\,dx/\theta=\theta/2.$$

Hence we can conveniently take the model,

$x_i=\theta/2+e_i$ where e_i is a random error.

Observed values (x_i)	1	3	6
Hypothetical values ($y_{i\theta}$)	$\theta/2$	$\theta/2$	$\theta/2$

A minimax estimate of θ is that value of θ for which $\max_i |x_i-\theta/2|$ is a minimum with respect to θ. Since all the observations should be between 0 and θ, it may be considered that the true value of θ is greater than or equal to 6. Let us examine $\max_i |x_i-\theta/2|$ for values of $\theta \geqslant 6$.

θ	6	7	8	9
$\max \mid x_i-\theta/2 \mid$	3	2.5	3	3.5

When $\theta=7$, $\max_i |x_i-\theta/2|$ is a minimum and hence $\hat{\theta}=7$.

Comments. It can be verified that for θ between 6 and 7 and 7 and 8, $\max_i |x_i-\theta/2|$ is greater than 2.5. The least square estimate of θ is easily seen to be 6.67 and if we minimize the dispersion D_1 for $r=1$ (see section 12.1) the estimate is 6. The maximum likelihood estimate is also 6. Due to lack of mathematical elegance the minimax procedure becomes difficult. A reader who finds some difficulty in following the arguments in this section

may omit this section. For other interpretations of the minimax procedures, such as treating the procedure as a minimization of the maximum risk when a decision is taken (an estimator is selected) in an unknown situation (for a parametric function) and for further reading see the references at the end of this chapter. Sometimes apriori probability distributions are assumed for the parameters in the population Such procedures are called Bayes' procedures (that is, statistical inference when the parameters are assumed to have prior distributions).

Exercises

12.6. Obtain the regression of Y on X from the following distributions :—

(1) $f(x, y) = \begin{cases} c.x(y+2) \text{ for } 0<x<1,\ 0<y<1 \\ 0 \text{ elsewhere and } c \text{ is a constant.} \end{cases}$

(2) $f(x, y) = \begin{cases} ax+2xy \text{ for } 0<x<1,\ 0<y<1 \\ 0 \text{ elsewhere and } a \text{ is a constant.} \end{cases}$

12.7. Obtain the regressions of Y on X and X on Y, given that

$$f(x \mid y) = \begin{cases} k\, y\, e^{-xy} \text{ for } x>0, \\ 0 \text{ elsewhere} \end{cases} \quad g(y) = \begin{cases} \tfrac{1}{2},\ 0<y<2 \\ 0 \text{ elsewhere} \end{cases}$$

where k is a constant.

12 8. If there is linear regression of X_1 on X_2 and X_3, (that is, $E(X_1 \mid X_2, X_3) = a+b\,x_2+c\,x_3$) (1) evaluate the regression coefficients, (2) write down the regression coefficients in terms of the various correlation coefficients and variances.

12.9. Assuming that there is linear regression of the marks obtained by the students, on the time spent, estimate the regression equation by the method of least squares from the following data.

Marks obtained (x)	50	555	60	67	75	80	82	88	90	95
Time spent (y)	2	25	3	3.5	4	4.4	4.6	4.8	5	5.1

Also obtain the sample correlation coefficient.

12.10. Assuming that there is linear regression of the marks obtained by the students on the time spent and the I.Q's of the students, estimate the regression equation by the method of least squares from the following data.

Marks obtained (x)	65	66	70	75	85	82	86	90
Time spent (y)	2	3	3.1	3.2	3	2.5	2.2	1.8
I.Q's (z)	100	98	101	102	105	103	110	1:0

12.11. If the relationship between x and y is assumed to be of the form $x=a+b(y-\bar{y})+e$, estimate a and b from the data given in problem 12.9, where e is a random error and $\bar{y}$ is the arithmetic mean of the observations on y.

12 12. If $x_1, x_2, \ldots x_n$ is an observed random sample from a $N(\mu=2, \sigma)$ obtain the least square estimate of σ^2 by constructing an appropriate model for σ^2.

[**Hint.** $(x_i-\mu)^2=\sigma^2+e_i$.]

12.13. If $x_1, x_2, \ldots x_n$ is an observed sample from an exponential population with parameter θ, estimate θ by minimizing the dispersion (1) D_1 for $r=1$, (2) D_1 for $r=2$ (see section 12.1).

[**Hint.** $EX_i=\theta$ for $i=1, 2, \ldots n$; $X_i=\theta+e_i$.]

12·14. If 5, 8, 10 is an observed random sample from an exponential population with parameter θ, estimate θ by minimizing the maximum dispersion [in the sense by minimizing D_2, (see Section 12.1)].

12.15. Assuming that e_i in a regression model $y_i=ax_i+b+e_i$, are independently distributed as a $N(0, \sigma)$ for $i=1, 2, \ldots, n$, obtain the maximum likelihood estimates of a and b. Show that they coincide with the least square estimates.

[**Hint.** $e_i : N(0, \sigma)$.

$$f(e_i)=\frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(y_i-ax_i-b)^2}{2\sigma^2}}=f(y_i \mid x_i)$$

$$L=\prod_{i=1}^{n} f(e_i)=\frac{1}{(\sqrt{2\pi})^n\sigma^n}\, e^{-\sum_{i=1}^{n}(y_i-ax_i-b)^2/2\sigma^2}$$

12.16. Show that the maximum likelihood estimates of σ^2 in problem 12.14 is

$$\hat{\sigma}^2=\sum_{i=1}^{n}[y_i-\hat{a}x_1-\hat{b}]^2/n$$

where ($\wedge$) denotes an estimated value. Express $\hat{\sigma}^2$ in terms of the sample correlation coefficient r.

12.17. Show that $\hat{a}$ and $\hat{b}$ are unbiased estimates for a and b in problem 12.14.

[**Hint.** $\Sigma(x_i-\bar{x})(y_i-\bar{y})=[\Sigma(x_i-\bar{x})(a(x_i-\bar{x})+e_i-e./n]=a.\Sigma(x_i-\bar{x})^2$
$+\Sigma(x_i-\bar{x})(e_i-e./n]$; $E[\Sigma(x_i-\bar{x})(y_i-\bar{y})=a.E\Sigma(x_i-\bar{x})^2=a.\Sigma(x_i-\bar{x})^2$ since the x's are constants].

12.18. Show that the variance of $\hat{a}$ in problem 12.14 is

$$\text{Var}(\hat{a})=\sigma^2/\Sigma_{i=1}^{n}(x_i-\bar{x})^2.$$

[**Hint** $E(e_i)=0$ for all i implies that $E(Y_i)=ax_i+b$ for all i ; $\text{Var}(e_i)=\sigma^2$ implies that $\text{Var}(Y_i)=\sigma^2$ for all i ; x's are constants.]

12.19. By using the results in problem 12.15 show that the null hypothesis $H_0 : \alpha=0$ (no linear regression) is equivalent to $H_0 : \rho=0$ where ρ is the linear correlation between Y and X.

12.6. EXPERIMENTAL DESIGNS

Here we will consider another problem where we use the principle of minimum dispersion, especially the principle of least squares, for estimating the parameters and the minimum value of dispersion, especially the least square minimum, is used for testing some hypotheses. When an experimenter finds it difficult to obtain the inference he had expected from his data, he often consults a statistician regarding this matter. In such cases usually it is seen that no valid conclusion can be drawn from the data, because the data was not collected properly or because the experiment was not conducted, keeping in mind the method of analysis of the experimental data. For example, suppose that an experimenter wants to compare two methods of teaching. Suppose that he conducts an experiment of teaching two classes of students according to the two different methods. The effect of a particular method of teaching is a hypothetical quantity. It cannot be measured directly. So he may take the average marks obtained by the students in the two classes. The average marks is not only a measure of the effect of a particular method of teaching but is also influenced by the different components of variation like, the intelligence of the students in the class, their previous aquaintance with the method of teaching, the particular teacher involved in the teaching etc. In this simple experiment there are so many extraneous factors which contribute to the average marks of the students. Hence a difference in the average marks cannot be taken as a measure of the difference in the effects of the two types of teaching. If an experimenter wants to compare two characteristics, his experiment should be designed in such a way that the data collected should in some sense measure only the characteristic (treatment) under study. A properly planned experiment controls the variation due to all other factors except the factors (treatments) under study. The reader might have noticed the necessity of properly designing an experiment in order to draw valid conclusions from the data. Here we will not consider the problem of how to design an experiment so that a particular analysis can be carried out. We will consider the analysis of the data, collected from a properly designed experiment.

12.61. One-way classification. Suppose that we want to compare the effects of 3 different types of fertilizers on the yield of corn. Suppose that 5 test plots which are homogeneous in all the extraneous variations such as fertility of the soil, climatic conditions etc., are planted with a variety of corn and fertilizer number one (say M_1) is applied in equal quantities. Another 5 plots which are also homogeneous in every respect are planted

with the same variety of corn and fertilizer number two (M_2) is applied in the same quantity as M_1. Similarly another set 4 of 5 plots are planted with the same variety of corn and M_3 is applied. Suppose that the yields of corn of these plots are as shown below.

M_1	M_2	M_3
13	8	9
10	9	9
10	10	10
8	8	8
9	7	12
50	42	48

The average yield of corn under the three fertilizers M_1, M_2, M_3 are 10, 8.4 and 9.6 respectively. We would like to know whether the observed differences in the average yields can possibly be attributed to chance, alone, or if instead the three fertilizers cannot be considered to be equally effective. We would like to test the hypothesis that there is no difference among the effects of the three fertilizers in the yield of this variety of corn. If the hypothesis is rejected, (that is, if the differences in the average yields cannot be taken as chance variation) we would like to test further whether M_1 is more effective than M_2, M_1 is more effective than M_3 etc., and also to estimate the difference in the effects of the fertilizers in terms of the observable quantity—the yield of corn. These are some of the problems that interest an experimenter in such an experiment. This problem reduces to the problem of testing the equality of means in three populations. So we will consider a general model for the analysis. In our example we have three sets of data, namely, 5 observations in each set.

Let
$$x_{ij} = \mu_i + e_{ij} \tag{12.38}$$

where x_{ij} denotes the j^{th} observation in the i^{th} set, μ_i is the effect of fertilizer number i and e_{ij} is a random error or an error due to chance variation in the j^{th} observation of the i^{th} set. In our example there are three fertilizers and 5 observations in each set and hence $i=1, 2, 3$ and $j=1, 2, 3 \ldots 5$. e_{ij}'s may be assumed to be independently and identically distributed. If the corn was planted in those test plots without the fertilizers M_1, M_2 and M_3 there would be some yields. So we shall take a more general model for our experiment

$$\mu_i = \mu + \alpha_i \tag{12.39}$$

where μ denotes a general effect and α_i denotes the deviation from the general effect due to fertilizer number i.

i.e.,
$$x_{ij} = \mu + \alpha_i + e_{ij}. \tag{12.40}$$

Without loss of generality we may assume that $\Sigma\ \alpha_i=0$. In general, if we have k treatments and n observations in each set, we can give the general model as

$$x_{ij}=\mu+\alpha_i+e_{ij} \qquad \text{for } i=1,2,\ldots k \qquad (12.41)$$
$$j=2,1,\ldots n$$

and e_{ij}'s are independently and identically distributed with mean zero and variance σ^2 ; $\mu, \alpha_1, \alpha_2, \ldots \alpha_k$ are all constants. If an experiment meets all the assumptions in the above model (12.41) and if the j^{th} observation in the i^{th} set can be expressed in the form $x_{ij}=\mu+\alpha_i+e_{ij}$ then such an experiment is said to be properly planned in order to use the model (12.41) for the analysis of the data. The model (12.41) is called a one-way classification experimental design model and the corresponding experimental data is said to belong to a one-way classification. Under the model (12.41) the hypothesis that there is no difference in the effects, that is,

$$H_0 : \mu_1=\mu_2=\ldots=\mu_k \text{ which is the same as}$$

$$H_0 : \alpha_1=0=\alpha_2=\ldots=\alpha_k, \text{ since } \Sigma\ \alpha_i=0 \text{ is assumed.}$$

Let us consider the estimation of the effects and testing of various hypotheses in a one-way classification. For convenience we will apply the principle of least squares in order to estimate the various effects.

$$e_{ij}=x_{ij}-\mu-\alpha_i \qquad (12.42)$$

$$e^2_{ij}=(x_{ij}-\mu-\alpha_i)^2 \qquad (12.43)$$

$$\sum_{i,j} e^2_{ij} = \sum_{i,j} (x_{ij}-\mu-\alpha_i)^2=\text{L (say)} \qquad (12.44)$$

where $$\sum_{i,j} = \sum_{i=1}^{k}\sum_{j=1}^{n} \qquad (12.45)$$

Minimize L with respect to the parameters $\mu, \alpha_1, \alpha_2, \ldots \alpha_k$.

It is easier to take $\mu_i=\mu+\alpha_i$ and minimize L with respect to $\mu_1, \mu_2, \ldots \mu_k$. The normal equations are

$$\frac{\partial L}{\partial \mu_1}=0,\ \frac{\partial L}{\partial \mu_2}=0,\ldots \frac{\partial L}{\partial \mu_k}=0 \qquad (12.46)$$

$$\frac{\partial L}{\partial \mu_i}=0 \Rightarrow -2\sum_{j=1}^{n}(x_{ij}-\mu_i)=0 \qquad (12.47)$$

($L=(x_{11}-\mu_1)^2+(x_{12}-\mu_1)^2+\ldots+(x_{1n}-\mu_1)^2+(x_{21}-\mu_2)^2+$
$+\ldots+\ldots+(x_{kn}-\mu_k)^2$ and differentiate L partially with respect to μ_i)

$$\Rightarrow \sum_{j=1}^{n} (x_{ij}-\mu_i)=0$$

$$\Rightarrow \sum_{j=1}^{n} (x_{ij})-\sum_{j=1}^{n} \mu_i=0$$

$$\Rightarrow x_{i\cdot}-n\mu_i=0$$

$$\therefore \qquad \hat{\mu}_i=\frac{x_{i\cdot}}{n} \tag{12.48}$$

where ($\wedge$) denotes an estimated value. We will use the notation

$$\sum_{i=1}^{k} x_{ij}=x_{\cdot j},\ \sum_{j=1}^{n} x_{ij}=x_{i\cdot}\ ,\ \sum_{i,\ j} x_{ij}=\sum_{i=1}^{k}\ \sum_{j=1}^{n} x_{ij}=x, \text{ etc.,}$$

that is, a summation with respect to a suffix is denoted by a dot.

But $$\mu_i=\mu+\alpha_i$$

$$\therefore \quad \hat{\alpha}_p-\hat{\alpha}_q=\hat{\mu}_p-\hat{\mu}_q=\frac{x_{p\cdot}}{n}-\frac{x_{q\cdot}}{n} \tag{12.49}$$

i.e., the difference in p^{th} and the q^{th} treaement effects is estimated by

$$\frac{x_{p\cdot}}{n} \quad \frac{x_{q\cdot}}{n}$$

The least square minimum S^2 or the minimum value of L is obtained by substituting $\hat{\mu}_i$ for μ_i, since $\mu_i=\hat{\mu}_i$ can be shown to minimize L.

$$S^2=\sum_{ij}(x_{ij}-\hat{\mu}_i)^2 \tag{12.50}$$

$$=\sum_{ij}\left(x_{ij}-\frac{x_{i\cdot}}{n}\right)^2=\sum_{tj} x_{ij}^2-\sum_{i}\frac{x_i^2}{n} \tag{12.51}$$

Let the data or the observations be as shown below :

	Set 1	*Set 2*	*Set 3*		*Set k*	
	x_{11}	x_{21}	x_{31}	....	x_{k1}	
	x_{12}	x_{22}	x_{32}	....	x_{k2}	
	x_{13}	x_{23}	x_{33}	....	x_{k3}	
	$\vdots$	$\vdots$	$\vdots$		$\vdots$	
	x_{1n}	x_{2n}	x_{3n}	....	x_{kn}	
Sum	$x_{1\cdot}$	$x_{2\cdot}$	$x_{3\cdot}$		$x_{k\cdot}$	$x_{\cdot\cdot}$

$\dfrac{x_{i\cdot}}{n}$ = the arithmetic mean in the i^{th} set

$\dfrac{x_{1\cdot}}{n}$ = arithmetic mean in the first set.

$\dfrac{x_{2\cdot}}{n}$ = arithmetic mean in the second set etc.

Hence $S^2 = \sum\limits_{i=1}^{k} \sum\limits_{j=1}^{n} \left(x_{ij} - \dfrac{x_{i\cdot}}{n}\right)^2$

may be called a measure of the within set variation.

Consider the hypothesis,

$$H_0 : \mu_1 = \mu_2 = \ldots\ldots = \mu_k$$

or

$$\alpha_1 = 0 = \alpha_2 = \ldots\ldots = \alpha_k \tag{12.52}$$

i.e., our hypothesis is that there is no difference among the treatment effects or the differences among the observed values may be attributed to chance variation. Under this hypothesis the model is

$$x_{ij} = \mu + 0 + e_{ij} \text{ for } i = 1, 2, \ldots k \quad j = 1, 2, \ldots n \tag{12.53}$$

The least square estimate of μ under H_0 is obtained by minimizing

$$L_0 = \sum_{ij} (x_{ij} - \mu)^2 \tag{12.54}$$

The normal equation is $\dfrac{\partial L_0}{\partial \mu} = 0$ (12.55)

$$\Rightarrow \sum_{ij} x_{ij} - \sum_{ij} \mu = 0$$

$$\Rightarrow \hat{\mu} = \frac{x_{\cdot\cdot}}{nk} \tag{12.56}$$

The least square minimum S_0^2 under H_0 is

$$S_0^2 = \sum_{ij} \left(x_{ij} - \frac{x_{\cdot\cdot}}{nk}\right)^2 \tag{12.57}$$

Since $\dfrac{x_{\cdot\cdot}}{nk}$ is the arithmetic mean of all the observations and S_0^2 / nk is the variance in all the observations, S_0^2 may be called a

measure of the total variation in the data. Hence $S_0^2 - S^2$ may be called a measure of the variation due to the hypothesis H_0. The minimum variation under the general model is S^2 and the minimum variation under the restricted model (restricted by H_0) is S_0^2. Therefore $S_0^2 - S^2$ can be attributed to the variation due to H_0.

$$S_0^2 - S^2 = \sum_{ij}\left(x_{ij} - \frac{x_{..}}{nk}\right)^2 - \sum_{ij}\left(x_{ij} - \frac{x_{i.}}{n}\right)^2$$

$$= \left(\sum_{ij} x_{ij}^2 - \frac{x^2_{..}}{nk}\right) - \left(\sum_{ij} x_{ij}^2 - \sum_i \frac{x_{i.}^2}{n}\right)$$

$$= \sum_i \frac{x^2_{i.}}{n} - \frac{x^2_{..}}{nk}$$

$$= n\sum_i\left(\frac{x_{i.}}{n} - \frac{x_{..}}{nk}\right)^2 \tag{12.58}$$

But $\frac{x_{i.}}{n}$ is the mean of the i^{th} set and $\frac{x_{..}}{nk}$ is the grand mean. Hence

$$S_0^2 - S^2 = n\sum_i\left(\frac{x_{i.}}{n} - \frac{x_{..}}{nk}\right)^2$$

may be called a measure of between set variation

$$S_0^2\left(= \sum_{ij} x_{ij} - \frac{x_{.}}{nk}\right)^2 = \sum_{ij} x^2_{ij} - \frac{x^2_{..}}{nk} = \text{total variation.} \tag{12.59}$$

$$S^2 = \sum_{ij}\left(x_{ij} - \frac{x_{i.}}{n}\right)^2 = \left(\sum_{ij} x^2_{ij} - \frac{x^2_{..}}{nk}\right) - \left(\sum_i x_{i.}^2 - \frac{x^2_{..}}{nk}\right)$$

$$= \text{within set variation.} \tag{12.60}$$

$$S_0^2 - S^2 = n\sum_i\left(\frac{x_{i.}}{n} - \frac{x_{..}}{nk}\right)^2 = \sum_i \frac{x^2_{i.}}{n} - \frac{x^2_{..}}{nk}$$

$$= \text{between set variation.} \tag{12.61}$$

But $S_0^2 = S^2 + \left(S_0^2 - S^2\right)$. (12.62)

Total variation = within set variation + between set variation.

The technique of splitting up the total variation in the data, into the variations due to the different components of variation, is called the analysis of variance technique. In our example we have taken account of the part of the total variation which can possibly be attributed to the variation due to the hypothesis $H_0 : \alpha_1 = 0 = \alpha_2 = \ldots = \alpha_k$, *viz*, that the different treatment effects are zero. We can test H_0 by testing the significance of the variation due to the null hypothesis $\left(\text{that is, } S_0^2 - S^2\right)$.

Our assumption in the model is that e_{ij}'s are independently and identically distributed with mean zero and variance σ^2. It may be easily shown that

$$\left(S_0^2 - S^2\right) \Big/ (k-1) \text{ and } S^2/(nk-k) = S^2/k(n-1)$$

are unbiased estimates of σ^2. (The proof of this part is left to the reader). Hence if we assume that e_{ij}'s are independently normally distributed with mean zero and variance σ^2 (this is equivalent to the statement that X_{ij}'s are independently normally distributed with mean $\mu + \alpha_i$ and variance σ^2 for $i=1, 2, \ldots k$ and $j=1, 2, \ldots n$) then

$$\frac{\left(S_0^2 - S^2\right)}{\sigma^2} = \frac{n}{\sigma^2} \sum_i \left(\frac{x_{i\cdot}}{n} - \frac{x_{\cdot\cdot}}{nk}\right)^2 : \chi^2_{k-1} \tag{12.63}$$

$$\frac{S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{ij} \left(x_{ij} - \frac{x_{i\cdot}}{n} \right)^2 : \chi^2_{k(n-1)} \tag{12.64}$$

But these two χ^2's can be shown to be independent. (The proof of this step is beyond the scope of this book).

$$\text{Hence } \quad \frac{\left(S_0^2 - S^2\right) \Big/ (k-1)}{S^2/k(n-1)} = F_{k-1,\, k(n-1)} \tag{12.65}$$

This variance ratio $F_{k-1,\, k(n-1)}$ may be tested for significance. If this F with $k-1$ and $k(n-1)$ degrees of freedom is not significant we can accept $H_0 : \alpha_1 = 0 = \ldots = \alpha_k$ or the treatment effects are all zero. If $F_{k-1,\, k(n-1)}$ is significant, not all the treatment effects are equal to zero. Some may be zero and some may not be zero. In this case we shall investigate further and see which are zero and which are not zero. This aspect will not be discussed in this book.

For convenience and simplicity the analysis may be given in a tabular form called the analysis of variance table.

Variation due to col. (1)	*Degrees of freedom (d.f.) col. (2)*	*Sum of squares (S.S.) col. (3)*	*Mean squares (M.S.) col. (3)/col. (2)*	*F-ratio*	*Inference*
Between sets (between samples)	$k-1$	$\left(S_0^2-S^2\right)$	$\left(S_0^2-S^2\right)/(k-1)=B$	B/E	
Within sets (error or residual)	$k(n-1)$	S^2	$S^2/k(n-1)=E$		
Total	$kn-1$	S_0^2			

To facilitate computation the following computational procedure may be given :—

1. Compute $\sum\limits_{ij} x_{ij}^2$ and $x_{..}$
2. Compute the correction factor (C.F.)$=x_{..}^2/nk$
3. Compute $\sum\limits_{i=1}^{k} \dfrac{x_{i.}^2}{n}$
4. Compute $S_0^2=\sum\limits_{ij} x_{ij}^2-\text{C.F.}$
5. Compute $S_0^2-S^2=\sum\limits_{i=1}^{k}\dfrac{x_{i.}^2}{n}-\text{C.F.}$
6. Obtain S^2 by subtraction $\left[S^2=S_0^2-\left(S_0^2-S^2\right)\right]$

After obtaining these the rest of the quantities in the analysis of variance table can be calculated easily.

Ex. 12.61.1. *The marks obtained by 10 students according to 3 different methods of teaching are given in the following table. Assuming that the experiment is planned in such a way that a one-way classification model may be set up for the data, test whether the three methods are equally effective, at 5% level.*

											Total
Method A	*50*	*60*	*60*	*65*	*70*	*80*	*75*	*80*	*85*	*75*	*700*
Method B	*60*	*60*	*65*	*70*	*75*	*80*	*70*	*75*	*85*	*80*	*720*
Method C	*40*	*50*	*50*	*60*	*60*	*60*	*65*	*75*	*70*	*70*	*600*

Sol. Let x_{ij} be the j^{th} observation in the i^{th} set

$$i=1, 2, 3 \text{ and}$$
$$j=1, 2, \ldots 10.$$

According to our notation $n=10$ and $k=3$.

$$x_{..}=\sum_{ij} x_{ij}=700+720+600=2020$$

$$\text{C.F.} = x_{..}^2 \Big/ nk = 136013.33$$

$$\sum_{i=1}^{n} \frac{x_{i.}^2}{n} = \frac{1}{10}[700^2+720^2+600^2]=136840$$

$$\sum_{i=1}^{n} \frac{x_{i.}^2}{n} - \text{C.F.} = 826.67$$

$$\sum_{ij} x_{ij}^2 - \text{C.F.} = 50^2+60^2+60^2+\ldots+70^2-\text{C.F.}=3636.67$$

The analysis of variance table.

Variation due to	*d. f.*	*S-S.*	*M. S.*	*F-ratio*	*Inference*
Between methods	$k-1=2$	826.67	413.335	$\frac{413.335}{1405.} < 1$	not significant
error	27	2810.00	1405.000		
Total	$nk-1=29$	3636.67			

Error sum of squares $=(3636.67)-(826.67)=2810.00$ and the corresponding degrees of freedom $=29-2=27$.

The tabled value of $F_{2,27}$ is in between 3.32 and 3.37 at the 5% level. The observed $F_{2,27}=413.335/1405<1$ and hence the hypothesis may be accepted or it may be assumed that the three methods of teaching are equally effective.

Comments. In this example and in the theory discussed, we assumed that the number of observations in each set is the same for all the sets. When the different sets have number of obervations $n_1, n_2, \ldots n_k$ where all the n_i's are not equal the theory may be developed in a similar fashion. Some exercises are given at the end of the chapter.

12.62. Two-way Classification. In our example of Section 12.61 we had three different fertilizers and one variety of corn

and we wanted to study the effect of these fertilizers on the yield of corn. Suppose that we want to study the effect of 3 different fertilizers on 4 different varieties of corn we can either conduct a number of experiments corresponding to each fertilizer variety combination or conduct a single experiment taking into account all the fertilizer variety combinations. In a planned experiment if we control the variations due to all the factors except the treatments, fertilizers and varieties and if x_{ij} denotes the observation corresponding to the i^{th} fertilizer and j^{th} variety, we can assume that x_{ij} is the result of a general effect, the effect due to the i^{th} fertilizer (say α_i), the effect due to the j^{th} variety (say β_j) and a chance variation (say e_{ij})

$$x_{ij}=\phi(\mu,\ \alpha_i,\ \beta_j,\ e_{ij}) \tag{12.66}$$

where ϕ is a function of (general effect), α_i, β_j and e_{ij}.

If we assume that $\phi=\mu+\alpha_i+\beta_j+e_{ij}$ (that is, the contribution of these different components is additive) we may write

$$x_{ij}=\mu+\alpha_i+\beta_j+e_{ij}. \tag{12.67}$$

In our example $i=1, 2, 3$ and $j=1, 2, 3, 4$ (since there are 3 fertilizers and 4 varieties). In general we may have r fertilizers and t varieties $i=1,2,\ldots r$ and $j=1,2,\ldots t$. In an experiment designed to study the effects of two types of treatments (for example fertilizers and varieties as described above) say $A_1, A_2,\ldots A_r$ and $B_1, B_2,\ldots B_t$ an observation corresponding to the i^{th} treatment of one type (A_i) and the j^{th} treatment of the other type (B_j) may be denoted by x_{ij}. If we have more than one observation corresponding to the (ij) combination of treatments we can denote the k^{th} observation corresponding to the (ij) treatment combination by x_{ijk}.

In the following discussion we shall consider the analysis when there is only a single observation corresponding to a treatment combination. A model of the form (12.68) is called a simple two-way classification model. If an experiment is designed in such a way that it satisfies all the conditions in the model (12 68) then such experimental data is said to belong to a simple two-way classification.

Model :
$$x_{ij}=\mu+\alpha_i+\beta_j+e_{ij} \qquad \begin{array}{l} i=1,2,\ldots r \\ j=1,2,\ldots t \end{array} \tag{12.68}$$

$\sum\limits_{i=1}^{r}\alpha_i=0$, $\sum\limits_{j=1}^{t}\beta_j=0$ and e_{ij}'s are independently and identically distributed with zero mean and variance equal to σ^2. Further $\mu, \alpha_1, \alpha_2,\ldots \alpha_r\ \beta_1, \beta_2,\ldots \beta_t$ are assumed to be constants. Such a model is called a simple, additive, fixed effect, two-way

classification model without interaction. The concept of interaction will be discussed later. The conditions $\Sigma\, \alpha_i=0=\Sigma\, \beta_j$ are justified since we can assume that α_i is the deviation from the general effect due to the factor A_i and β_j is the deviation from the general effect due to the factor (or treatment) B_j.

12.62.1. Estimation of the parameters. The parameters μ, α_i, β_j, for $i=1, 2,\ldots r$ and $j=1, 2\ldots t$ may be easily estimated by the principle of least squares

$$e_{ij}=x_{ij}-\mu-\alpha_i-\beta_j \tag{12.69}$$

$$\sum_{i=1}^{r} \sum_{j=1}^{t} e_{ij}^2=\sum_{ij} e_{ij}^2=\sum_{ij} (x_{ij}-\mu-\alpha_i-\beta_j)^2=\text{L (say)}. \tag{12.70}$$

Minimize L with respect to μ, $\alpha_1,\ldots \alpha_r$, $\beta_1,\ldots\beta_t$.

For convenience let $\mu_i=\mu+\alpha_i$.

$\therefore$ The normal equations are $\dfrac{\partial L}{\partial \mu_i}=0$, for $i=1, 2,\ldots r$

and $$\frac{\partial L}{\partial \beta_j}=0 \text{ for } j=1, 2,\ldots t. \tag{12.71}$$

$$\frac{\partial L}{\delta \mu_i}=0, \Rightarrow -2.\sum_{j=1}^{t} (x_{ij}-\mu_i-\beta_j)=0.$$

$$\Rightarrow \sum_{j=1}^{t} (x_{ij}-\mu_i-\beta_j)=0 \tag{12.72}$$

$$\Rightarrow x_{i\cdot}-t\,\mu_i=0. \qquad \left(\text{since } \sum_{j=1}^{t} \beta_j=0\right)$$

$$\therefore \quad \hat{\mu}_i=\frac{x_{i\cdot}}{t} \tag{12.73}$$

$$\therefore \quad \hat{\alpha}_p-\hat{\alpha}_q=\hat{\mu}_p-\hat{\mu}_q=\frac{x_{p\cdot}}{t}-\frac{x_{q\cdot}}{t} \tag{12.74}$$

$$\frac{\partial L}{\partial \beta_j}=0, \Rightarrow \sum_{i=1}^{r} (x_{ij}-\mu_i-\beta_j)=0$$

$$\beta_j=\frac{x_{\cdot j}}{r}=\frac{x_{\cdot\cdot}}{rt} \tag{12.75}$$

$$\beta_p-\hat{\beta}_q=\frac{x_{\cdot p}}{r}-\frac{x_{\cdot q}}{r} \tag{12.76}$$

The least square minimum

$$S^2 = \left(\sum_{ij} x_{ij} - \frac{x_{i.}}{t} - \frac{x_{.j}}{r} + \frac{x_{..}}{rt} \right)^2$$

(since $\mu_i = \hat{\mu}_i$ and $\beta_j = \hat{\beta}_j$ can be shown to minimize L)

$$S^2 = \left(\sum_{ij} x_{ij}^2 - \frac{x^2_{..}}{rt} \right) - \left(\sum_i \frac{x_{i.}^2}{t} - \frac{x_{..}}{rt} \right) - \left(\sum_j \frac{x^2_{.j}}{r} - \frac{x^2_{..}}{rt} \right)$$

(12.77)

The simplification is left to the reader. The fact that $\sum_{ij} (\hat{\mu}_i + \hat{\beta}_j)(x_{ij} - \hat{\mu}_i - \hat{\beta}_j) = 0$, due to the normal equations, may be used to advantage)

Let us consider the hypothesis that there is no difference in the effects of $A_1, A_2, \ldots A_r$, that is,

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_r = 0$$

or

$$\mu_1 = \mu_2 = \ldots = \mu_r = \mu. \qquad (12.78)$$

Under H_0, the model becomes

$$x_{ij} = \mu + \beta_j + e_{ij}. \qquad (12.79)$$

The least square estimates are

$$\hat{\mu} = \frac{x_{..}}{rt} \text{ and } \hat{\beta}_j = \frac{x_{.j}}{e} - \frac{x_{..}}{rt} \qquad (12.80)$$

and the least square minimum S_0^2 under H_0 is

$$S_0^2 = \left(\sum_{ij} x_{ij}^2 - \frac{x^2_{..}}{rt} \right) - \left(\sum_j \frac{x^2_{.j}}{r} - \frac{x^2_{..}}{rt} \right). \qquad (12.81)$$

$\therefore$ The sum of squares due to the hypothesis H_0 is

$$S_0^2 - S^2 = \sum_i \frac{x_{i.}^2}{t} - \frac{x^2_{..}}{rt}$$

This may be called the sum of squares due to the A factors or the treatments $A_1, A_2, \ldots A_r$. Similarly the sum of squares due to be B factors may be obtained as

$$\sum_j \frac{x_{.j}^2}{r} - \frac{x^2_{.}}{rt} \qquad (12.81)$$

The analysis of variance table for testing these two hypotheses may be set up as follows.

Variation due to	*d.f.*	*S. S.*	*M.S.*	*F—ratio*	*Inference*
A—treatments	$r-1$	$\sum_i \frac{x_{i\cdot}^2}{t} - \frac{x^2_{\cdot\cdot}}{rt} = T_1$	$T=T_1/(r-1)$	T/E	
B—treatments	$t-1$	$\sum_j \frac{x^2_{\cdot j}}{r} - \frac{x^2_{\cdot\cdot}}{rt} = T_2$	$U=T_2/(t-1)$	U/E	
Error	$rt-r-t+1$ $=(r-1)(t-1)$	(subtraction)$=E_1$	$E=E_1/(n-1)(t-1)$		
Total	$rt-1$	$\sum_{ij} x_{ij}^2 - \frac{x^2_{\cdot\cdot}}{rt} = T_3$			

If we assume that e_{ij}'s are independently normally distributed with parameters $\mu=0$ and σ, then by an argument similar to the one used in a one-way classification analysis, we can show that T/E is an F with $r-1$ and $(r-1)(t-1)$ degrees of freedom, and U/E is an F with $t-1$ and $(r-1)(t-1)$ degrees of freedom. Hence the hypotheses $\alpha_1=\alpha_2=\ldots=\alpha_r=0$ and $\beta_1=\beta_2=\ldots=\beta_t=0$ may be tested by T/E and U/E respectively.

A two-way classification of data is given in the following table :—

	B_1	B_2	....	B_t	Total
A_1	x_{11}	x_{12}		x_{1t}	$x_{1\cdot}$
A_2	x_{21}	x_{22}		x_{2t}	$x_{2\cdot}$
.					
.					
.					
A_r	x_{r1}	x_{r2}		$x_{\cdot t}$	$x_{r\cdot}$
Total	$x_{\cdot 1}$	$x_{\cdot 2}$		$x_{\cdot t}$	$x_{\cdot\cdot}$

Computational procedure :

1. Compute $\sum_{ij} x_{ij}^2$ and $x_{..}$

2. Compute the correction factor (C.F.)$=\dfrac{x_{..}^2}{rt}$

3. Compute $\sum_{i} \dfrac{x_{i.}^2}{t}$ —C.F.

4. Compute $\sum_{j} \dfrac{x_{.j}^2}{r}$ —C.F.

5. $d.f.$ for error$=(rt-1)-(r-1)-(t-1)=(r-1)(t-1)$

6. S.S. for error$=T_3-T_1-T_2$

Ex. 12.62.1. *The observations from an experiment which is designed to use the model (12.68) and compare the effects of 3 types of fertilizers on the yields of 4 varieties of corn, are given in the following table. Test the hypotheses (1) there is no difference among the fertilizers (2) There is no difference among the varieties, as far as the yields are concerned.*

Varieties / *Fertilizers*	B_1	B_2	B_3	B_4	*Total*
A_1	15	20	22	20	77
A_2	20	30	32	28	110
A_3	20	35	38	32	125
Total	55	85	92	80	312

Sol. C.F.$=(312)^2/12=8112$

$$\sum_{i} \frac{x_{i.}^2}{t}-\text{C.F.}=(77^2+110^2+125^2)/4-8112$$

$$=301{\cdot}5$$

$$\sum_j \frac{x^2_{\cdot j}}{r} - \text{C.F.} = (55^2+85^2+92^2+80^2)/3 - 8112$$

$$= 279.3$$

$$\sum_{ij} x^2_{ij} - \text{C.F.} = (15^2+20^2+\ldots+32^2) - 8112$$

$$= 598$$

The residual s.s. $= 598 - (279.3 + 301.5)$
$= 17.2$

Here $r=3$ and $t=4$. The various degrees of freedom are $r-1=2$, $t-1=3$, $(r-1)(t-1)=6$, $rt-1=11$. The analysis of variance table is,

Variation due to	*d.f.*	*S.S.*	*M.S.*	*F-ratio*	*Inference*
Between fertilizers	2	301.5	150.75	$\frac{150.75}{2.87}=52.9$	Significant
Between varieties	3	279.3	93.10	$\frac{93.10}{2.87}=32.4$	Significant
Residual	6	17.2	2.87		
Total	11	598.0			

The tabled values of F at the 1% level are

$F_{2,\,6}=10.9$ and $F_{3,\,6}=9.78$.

The observed F-values are greater than the tabled values and therefore we reject the hypothesis. We can not assume that the varieties are the same or that the fertilizers are the same, as far as the yields are concerned.

Comments. Individual hypothesis such as whether any two particular varieties have equal effects or any two fertilizers have equal effects, etc., may be tested by using a student t test. This will not be discussed in this book.

A two-way classification is different from a contingency table. In a contingency table we have frequencies whereas in a two-way classification table we have observations.

In the example discussed above we have considered the problem of taking single observation corresponding to every treatment combination. For example corresponding to the treatments A_1 and B_1 we take one observation etc. Instead of taking single observation if we have taken n observations corresponding to every treatment combination, the analysis can be carried out in a similar fashion. The observations corresponding to A_i and B_j may be written as $x_{ij1}, x_{ij2}, \ldots x_{ijn}$. If the number of observations in each cell is not the same or if there are n_{ij} observations corresponding to A_i and B_j for $i=1, 2, \ldots r$ and $j=1, 2, \ldots t$ (that is, n_{11} observations in $(1, 1)^{th}$ cell, n_{12} observations in the $(1, 2)^{th}$ cell etc.) then the normal equations and the analysis become complicated.

In the model (12.68) we have taken only the variation due to A_i and B_j. There is a possibility of a joint variation in the sense that there may be a contribution due to the particular combination (A_i, B_j) or the effect of A_i may be varying with respect to the B factors. In such cases the joint variation is called interaction and the model (12.68) may be modified as,

$$x_{ij}=\mu+\alpha_i+\beta_j+\gamma_{ij}+e_{ij} \tag{12.84}$$

where γ_{ij} denotes the interaction between A_i and B_j. Study of interaction is not possible in the model (12.84) unless we have more than one observation corresponding to A_i and B_j or we put more restrictions on the model (12·84). If we have a number of observations corresponding to A_i and B_j we can write the k^{th} observation in the $(ij)^{th}$ cell as

$$x_{ijk}=\mu+\alpha_i+\beta_j+\gamma_{ij}+e_{ijk} \tag{12.85}$$

and the analysis may be carried out. These ideas and definitions may be easily generalized for a p-way classification.

Sometimes we may have to take more than one set of observations corresponding to a treatment in order to have a valid inference. For example consider an experiment in which k different diets are tried on K sets of a particular variety of cattle. If all types of extraneous variations are controlled a model of the form

$$x_{ij}=\mu+\alpha_i+e_{ij}$$

is in order, where x_{ij} denotes the increase in weight. But it is difficult to get animals of the same initial weight and the initial weights may have some effect on the increase in weight. The initial weights (say y_{ij}) are observable and hence we may use a modified model

$$x_{ij}=\mu+\alpha_i+by_{ij}+e_{ij} \tag{12.86}$$

where b is a constant. This additional variable is often called a concomitant variable. For the simplicity of the analysis we assumed that the concomitant variable is related to x_{ij} in the form of a linear regression as given in the model (12.86). The analysis of a model involving one or more concomitant variables is called the

analysis of covariance. If there is a concomitant variable present in a two-way classification model without interaction, we may use the modified model

$$x_{ij}=\mu+\alpha_i+\beta_j+cy_{ij}+e_{ij} \tag{12.87}$$

where c is a constant.

12.63. Randomized Block Experiments. This is a special experimental design where we can use a two-way classification model without intersection for the analysis of the experimental data. Consider an agricultural experiment, designed to study the effect of t types of fertilizers on the yield of wheat. Take a section of land called a block which is homogeneous in all the extraneous variations such as fertility of the soil, climatic conditions etc. Divide into t plots of the same shape and dimensions. Apply the t fertilizers at random to these t plots in order to avoid possible variations between plots due to uncontrolled factors. Plant the wheat of the particular variety under consideration. We shall replicate or repeat the experiment by taking another block

F_1	F_5	F_1	.	.	.	.

of land which is homogeneous within, dividing it into t plots and applying the fertilizers at random and planting the same type of wheat. These blocks are homogeneous within but there may be between block variations. For example one block may be in one province and the other may be in another province. If we have r such blocks and all the t treatments tried in all the blocks, we have a randomized block experiment. A convenient model for a randomized block experiment is

$$x_{ij}=\mu+\alpha_i+\beta_j+e_{ij} \tag{12.88}$$

α_i—block effects and β_j—treatment effects.

If we have a large number of treatments it may be difficult to get a homogeneous block so that we can try all the treatments. In such cases other designs known as incomplete block designs are used. Some of the commonly used incomplete block designs are : (1) balanced incomplete block design, (2) partially balanced incomplete block designs, (3) lattice designs etc.

12.64. Latin Square Designs. A latin square of order n is an arrangement of n elements into n rows and n columns such that every element appears in each row and column once and only once. A latin square of order 3 is given below. The elements are the latin letters A, B and C.

A	B	C
B	C	A
C	A	B

Such a design can be conveniently used to test hypotheses regarding three sets of treatments $A_1, A_2, \ldots A_n$; $B_1, B_2, \ldots B_n$ and $C_1, C_2, \ldots C_n$.

Take one set of treatments corresponding to the rows, another corresponding to the columns and the third corresponding to the latin letters. The

various hypotheses may be tested by taking [as small as n^2 (number of cells in the latin square) observations. A convenient fixed effect non-interaction model is

$$x_{ij(k)}=\mu+\alpha_i+\beta_j+\gamma_k+e_{ij(k)} \tag{12.89}$$

α_i's, β_j 's and γ_k's are the effects of the three sets of treatments. For convenience we will call α_i's the row effects, β_j's the column effects and γ_k's the treatment effects.

$$\sum_{i=1}^{n} \alpha_i=0=\sum_{j=1}^{n} \beta_j=\sum_{k=1}^{n} \gamma_k,\ E[e_{ij(k)}]=0 \text{ and } \operatorname{Var}[e_{ij(k)}]=\sigma^2$$

are assumed. $x_{ij(k)}$ denotes the observation corresponding to the treatment C_k if it appears in the $(ij)^{th}$ (i^{th} row j^{th} column) cell. If we assume further that $e_{ij(k)}$ are independently distributed as a $N(0, \sigma)$ then the analysis of variance table for a latin square experiment may be given as follows. (Here the normality assumption is used only to test the various hypotheses by using F-tests).

Variation due to	*d.f.*	*S.S.*	*M.S.*	*F-ratio*	*Inference*
Rows	$n-1$	$\sum_i \frac{x_{i..}^2}{n}-\frac{x_{...}^2}{n^2}$	(col. 3)/(col.2) =R	R/E	
Columns	$n-1$	$\sum \frac{x_{.j}^2}{n}-\frac{x_{...}^2}{n^2}$	C	C/E	
Treatments	$n-1$	$\sum_k \frac{x_{..k}^2}{k}-\frac{x_{...}^2}{n^2}$	T	T/E	
Residual	$(n-1)(n-2)$	(by subtraction)	E		
Total	(n^2-1)	$\sum_{ij} x_{ij}^2-\frac{x_{...}^2}{n^2}$			

The degrees of freedom for the residual

$$=(n^2-1)-(n-1)-(n-1)-(n-1).$$

The residual sum of squares = Total S.S. − Row S.S. − Col. S.S. − Treat S.S.

The various F-ratios in the 5th column have $n-1$ and $(n-1)(n-2)$ degrees of freedom. Other designs related to a latin square design are greeco-latin squares, youden square etc.

There are a number of other designs which are commonly used, such as factorial designs, split-pilot designs, nested designs etc. So far we have been considering only fixed effect models. That is, for example, if we have a one-way classification model,

$$x_{ij}=\mu+\alpha_i+e_{ij}$$

we assumed that α_i's are constants. If the treatments used are only a random sample from a finite or an infinite set of treatments, α_i's may be assumed to be the values assumed by a stochastic variable. A model, in which the various effects are assumed to be stochastic variables, is called a random effect or a variable effect model. In a model if some effects are assumed to be constants and some as variables such a model is called a mixed effect model. For a detailed discussion of the different designs and models see the bibliography at the end of this chapter.

Exercises

12.19. The following table gives the I.Q's of 4 groups of students. Assuming that these observations may be considered to be random samples from independent populations $N(\mu_i, \sigma)$ for $i=1, 2, 3, 4$, test the hypothesis that $\mu_1=\mu_2=\mu_3=\mu_4$ at the 5% level by using an analysis of variance technique.

Group 1. 100, 102, 105, 101, 98, 115, 112, 110, 114, 108.

Group 2. 98, 100, 100, 102, 95, 110, 108, 112, 111, 104.

Group 3. 105, 102, 103, 104, 99, 100, 107, 106, 107, 102.

Group 4. 99, 105, 110, 112, 98, 102, 98, 100, 100, 104.

12.20. The results of a completely randomized experiment (where a one-way classification model is appropriate) conducted to study the yields of 3 varieties of wheat are given below. Test at the 5% level whether all the varieties can be considered to be the same as far as the yields are concerned.

Yields

Variety 1.	10	12	11	13	15
2.	14	15	17	20	16
3.	12	14	15	13	11

12.21. In a one-way classification with the model

$x_{ij}=\mu+\alpha_i+e_{ij}$ $i=1, 2,\ldots k$; $j=1, 2,\ldots n$

$E(e_{ij})=0$ and Var $(e_{ij})=\sigma^2$ for all i and j

Show that

(1) $\sum_{ij} x_{ij}^2 - \frac{x^2_{..}}{nk} = \left(\sum_i \frac{x^2_{i.}}{n} - \frac{x^2_{..}}{nk} \right) + \left(\sum_{ij} \frac{x^2_{ij}}{n} - \frac{x^2_{..}}{nk} \right)$

(2) $E\left[\sum_{ij} x^2_{ij} - \frac{x^2_{..}}{nk} \right] = (nk-1)\,\sigma^2$;

$$E\left[\sum_i \frac{x^2_{i.}}{n} - \frac{x^2_{..}}{nk} \right] = (k-1)\,\sigma^2 :$$

$$E\left[\sum_{ij} x^2_{ij} - \sum_i \frac{x^2_{i.}}{n} \right] = k(n-1)\,\sigma^2$$

under the hypothesis $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$.

12.22. In a one-way classification, if the number of observations in the i^{th} set is n_i for $i=1, 2, \ldots k$, show that the sum of squares due to, between set variation is

$$\sum_i \frac{x^2_{i.}}{n_i} - \frac{x^2_{..}}{n.} \text{ where } x_{ij}, x_{i.}, x_{..} \text{ and } n.$$

denote the j^{th} observation in the i^{th} set, $\sum_j x_{ij} = x_{i.}$, $\sum_{ij} x_{ij} = x_{..}$ and $\sum_i n_i = n.$ respectively.

[**Hint.** Use the model $x_{ij} = \mu_i + e_{ij}$ for $i=1, 2, \ldots k$ and $j=1, 2, \ldots n_i$;

Obtain the least square minima under the general model and under the restricted model, restricted by the hypothesis $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k$.]

12.23. Show that the degrees of freedom for the residual in the model of problem 12.14 is $(n.-1)-(k-1)=(n.-k)$.

12.24. For the observations in problem 12.14 show that

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(x_{ij} - \frac{x_{..}}{n.} \right)^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} \left(x_{ij} - \frac{x_i}{n} \right)^2$$
$$+ \sum_{i=1}^{k} n_i \left(\frac{x_{i.}}{n_i} - \frac{x_{..}}{n.} \right)^2$$

12.25. Three processes A, B and C are used in a production process. Assuming that the design is such that we can use a one-way classification model, test at the 5% level whether the three processes can be considered to be equivalent as far as the outputs are concerned. The following observations on the outputs are made

Method A 10, 12, 13, 11, 10, 14, 15, 13

Method B 9, 11, 10, 12, 13

Method C 11 10, 15, 14, 12, 13

12.26. The yields of 3 varieties of corn according to two methods of planting, are given in the following table. Assuming that a two-way classification model without intersection is appropriate for the analysis, test the various hypotheses at the 5% level.

	Variety B_1	*Variety* B_2	*Variety* B_3
Method A_1	10	15	22
Method A_2	8	16	26

12.27. The results of a randomized block experiment conducted to study the effects of three methods of spinning on the breaking strength of 3 types of cotton, are given in the following table. The observations are the 'breaking strengths'. Test the hypothesis that the three methods are equally effective, at the 5% level.

	Type 1	*Type 2*	*Type 3*
Method 1	22	21	20
Method 2	18	20	24
Method 3	19	22	24

12.28. In a two-way classification model with interaction

$$x_{ijk}=\mu+\alpha_i+\beta_j+\gamma_{ij}+e_{ijk}\ ;\ i=1, 2,\ldots r\ ;$$

$j=1, 2,\ldots t$; $k=1, 2,\ldots n$; obtain the sum of squares due to interaction.

12.29. If $n=1$ in problem 12.28 can the interaction sum of squares be estimated ?

12.30. Set up the analysis of variance table for the model in problem 12.28.

12.31. In a latin square design with the model

$$x_{ij(k)}=\mu+\alpha_i+\beta_j+\gamma_k+e_{ij(k)}$$

$i=1, 2,\ldots n$; $j=1, 2,\ldots n$; $k=1, 2,\ldots n$; show that the residual sum of squares is

$$\left(\sum_{ijk} x^2_{ij(k)}-\frac{x^2_{\cdots}}{n^2}\right)-\left(\sum_i \frac{x^2_{i..}}{n}-\frac{x^2_{\cdots}}{n^2}\right)$$

$$-\left(\sum_j \frac{x^2_{.j.}}{n}-\frac{x^2_{\cdots}}{n^2}\right)-\left(\sum_k \frac{x^2_{..k}}{n}-\frac{x^2_{\cdots}}{n^2}\right)$$

where $x_{i..}$, $x_{.j.}$ and $x_{..k}$ denote the sum of the observations in the i^{th} row, j^{th} column and corresponding to the k^{th} treatment respectively.

12.32. Show that the degrees of freedom for the residual in problem 12.25 is $(n^2-1)-3(n-1)=(n-1)(n-2)$.

12 33. A latin square design and the corresponding observations are given below. Write down the analysis of variance table and test the various hypotheses at the 5% level.

Design

A	B	C
B	C	A
C	A	B

Observations

8	7	10
10	12	14
9	14	16

12.7. A GENERAL TEST STATISTIC

In the experimental design problems, in section 12.6 we used the variance-ratio tests for testing the various null hypotheses. The least square minima under the various hypotheses were obtained and under the assumption of normality for the error in the model, the comparison of the least square minima led to a variance-ratio test. In general the minimum dispersion may be used as a criterion for testing various hypotheses regarding a mathematical model set up for observed data. A comparison of the minimum value of a measure of dispersion will yield test statistics for testing various hypotheses.

12.71. Kolmogorov-Smirnov Statistic. This is a convenient statistic for testing 'goodness of fit' and was formulated by two Russian mathematicians and hence the statistic is named after them. In chapter 11 we discussed the use of a chi-square statistic for testing 'goodness of fit'. But in order to use the chi-square test the data had to be classified and further more the frequencies must be sufficiently large. This condition restricts the use of a chi-square statistic when the frequencies are small or when a classification is not desirable for a given data. In such a situation an exact test is usually given by the Kolmogorov-Smirnov statistic D_n.

Consider a goodness of fit problem. Let $x_1, \ldots, x_n$ be the observed data. We want to test whether this sample can be considered to be a random sample from a specified distribution $f(x, \theta_0)$. For example, suppose we have an observed sample of size 15 and that we would like to decide whether this sample has come from an exponential population with the parameter $\theta=2$. Let $u_1 \leqslant u_2 \leqslant \ldots \leqslant u_n$ be the arrangement of $x_1, \ldots, x_n$ according to

the order of the magnitude of the x's. Then the sample distribution function is given by the formula,

$$S_n(x)=\begin{cases}0 & x<u_1\\ r/n & u_r\leqslant x<u_{r+1}\\ 1 & x\geqslant u_n\end{cases} \tag{12.99}$$

$S_n(x)$ is the cumulative frequency function. Let the hypothesis to be tested, be that the x's came from a population with the density function $f(x, \theta_0)$ or with the distribution function $F(x, \theta_0)$ where θ_0 denotes that the parameters are fixed or that the distribution is completely specified. Therefore the hypothetical value for $S_n(x)$ is $F(x, \theta_0)$. The error in this model is

$$e=S_n(x)-F(x, \theta_0) \tag{12.100}$$

If we use the measure of dispersion D_5 (see section 12.1 equation 12.7) then

$$D(e)=\sup_x \mid S_n(x)-F(x, \theta_0) \mid \tag{12.101}$$

The minimum dispersion

$$=\min_\theta D(e)=\sup_x \mid S_n(x)-F(x, \theta_0) \mid$$

[since all the parameters are specified $\min\limits_\theta D(e)=D(e)$] (12.102)

$$D_n= \sup_x \mid S_n(x)-F(x, \theta_0) \mid \tag{12.103}$$

is called the Kolmogorov-Smirnov statistic. Instead of using the measure D_5, if we use the measure of dispersion D_4 for $r=2$ (see section 12.1, equation 12.6) we get

$$W=\min_\theta \{E_x \mid S_n(x)-F(x, \theta_0) \mid ^2\}^{1/2}$$

$$=\{E_x \mid S_n(x)-F(x, \theta_0) \mid ^2\}^{1/2} \tag{12.104}$$

W^2 is usually called W^2-statistic for goodness of fit. Evidently D_n and W^2 are stochastic variables. Their distributions can be worked out for any particular sample size n. The critical values are tabulated for various values of sample size and hence the 'goodness of fit' may be tested by using a D_n or W^2 statistics.

Ex. 12.71.1. *Use a D_n statistic to test whether a normal distribution with the parameters $\mu=13$ and $\sigma=1$ is a good fit for the data given below : 9, 10, 10, 11, 12, 12, 13, 13, 13, 14, 14, 15, 15, 16.*

Sol. $n=14$; $f(x, \theta_0)=\dfrac{1}{\sqrt{2\pi}}\, e^{-\frac{(x-13)^2}{2}}$

$$F(x, \theta_0) = \int_{-\infty}^{x} f(x, \theta_0)dx = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \text{ whele } t = \frac{x-13}{1}$$

$$S_n(x) = S_{14}(x)$$

The various values are given in the following table :

u	frequencies	$S_{14}(x)$	x	$F(x, \theta_0)$	$\mid S_{14}(x) - F(x, \theta_0) \mid$
			$x<y$	0.0000	0.0000
9	1	1/14=0.0714	$9 \leqslant x < 10$	0.0013	0.0701
10	2	3/14=0.2142	$10 \leqslant x < 11$	0·0228	0.1914
11	1	4/14=0.2856	$11 \leqslant x < 12$	0.1587	0.1269
12	2	6/14=0.4284	$12 \leqslant x < 13$	0.5000	0.0716
13	3	9/14=0,6426	$13 \leqslant x < 14$	0.8413	0.1987
14	2	11/14=0·7854	$14 \leqslant x < 15$	0.9772	0.1918
15	2	13/14=0.9282	$15 \leqslant x < 16$	0.9987	0.0705
16	1	14/14=1.0000	$x \geqslant 16$	0.9999	0.0001

$F(x, \theta_0)$ for various values of x are obtained from a normal probability table. For example, when $9 \leqslant x < 10$,

$$F(x, \theta_0) = \int_{-\infty}^{10} (2\pi)^{-1/2} e^{-(x-13)^2/2} dx$$

$$= \int_{-\infty}^{-3} (2\pi)^{-1/2} e^{-t^2/2} dt$$

$$= 0.0013.$$

$$D_{14} = \max \mid S_{14}(x) - F(x, \theta_0) \mid = 0\ 1987.$$

The tabulated value of D_{14} at the level $\alpha = 0.05$ is 0.35 approximately. This is obtained from a table of D_n. (References are given at the end of this chapter). The observed $D_n = 0.1798 < 0.35$ and hence a $N(\mu = 13, \sigma = 1)$ may be considered to be a good fit at the 5% level.

Comments. In most of the tests mentioned in this chapter we did not assume a basic distribution. Such tests are often

called distribution free tests. In D_n and W^2 statistics we did not consider a problem of testing a null hypothesis against a parametric alternative. These distribution-free tests are sometimes called non-parametric tests. Some other non-parametric tests commonly used, are sign test, rank test, run test etc. A brief account of the some of the non-parametric tests are given in the following section. For further reading refer to be bibliography given at the end of this chapter.

Exercises

12.34. Test the goodness of fit of a Poisson distribution with the parameter $\lambda=2$ to the following frequency table, at the 5% level, by using a Kolmogorov-Smirnov statistic.

x	0	1	2	3	4	5	6	7	8
Frequency	52	63	60	40	22	10	3	1	0

The tabled value of $D_{256}=0.085$.

12.35. Test the goodness of fit of a $N(\mu, \sigma)$ to the following frequency table by using a Kolmogorov-Smirnov statistic

x	10	12	13	14	15	16	17
Frequency	2	4	6	10	7	4	2

[**Hint.** Estimate the parameters. The test is not exact due to the estimation of the parameters. Use the 1% level. The tabled value at 1% level is $D_{35}=0.27$.]

12.8. SOME DISTRIBUTION-FREE PROCEDURES

So far we discussed statistical inference when the population under consideration was assumed to be normal in the case of continuous populations. Even though there are a good number of theoretical results which will justify normality assumption, there are situations where nothing is known about the underlying distribution or a normality assumption, may not be desirable. In such cases we resort to some distribution-free procedures, that is, procedures where a particular basic distribution is not assumed. In chapters 10 and 11 we were considering only parametric tests, that is, tests where a particular hypothesis is tested against a parametric alternative. In other words the tests were restrictions on estimable parametric functions. There are other situations such as testing the independence of populations, randomness of data, compatibility of a particular data to a theoretical distribution, departure from normality etc. A few distribution-free procedures will be considered here.

12.81. The Sign Test. This is a distribution-free test which can be conveniently applied when the underlying population is known to be continuous and symmetrical. If we want to test a hypothesis that a parameter $\theta=\theta_0$ where θ_0 is a specified value of θ and if we have a random sample which are observations on θ then a single sample sign test can be formulated as follows. Assign a plus sign to all the observations which exceed θ_0 and a minus sign to all the observations which are less than θ_0. If the population is known to be symmetrical about the ordinate at $\theta=\theta_0$ the probability p of getting a plus sign can be taken to be 1/2 without any loss of generality. So the hypothesis $\theta=\theta_0$ reduces to the hypothesis $p=1/2$ where p is the probability of a success in a Binomial probability situation. Since the population is continuous, strictly speaking, the probability of getting an observation equal to θ_0 is zero. If there is an observation equal to θ_0 this is caused by rounding of the numbers and hence it may be omitted. In this case the alternative to the hypothesis $p=1/2$ can be formulated as $p<1/2$, $p>1/2$, or $p\neq 1/2$ according to the requirement of the experimenter.

Ex. 12.81.1. *The breaking strength measured at random, of cotton thread spun by a particular process is given in the following data. Test the hypothesis that the expected breaking strength is 10 units, at the 5% level. Assume that the population is continuous and symmetrical and a sign test is in order. 10.1, 10.3, 10.5, 10.1, 10.0, 9.7, 9.8, 9.7, 9.9, 10.2, 10.2, 10.4, 10.1, 9.9, 9.8, 10.1, 10.3.*

Sol. We want to test $\left.\begin{matrix} H_0 : \theta=10 \\ H_1 : \theta\neq 10 \end{matrix}\right\} \Rightarrow \left.\begin{matrix} H_0 : p=1/2 \\ H_1 : p\neq 1/2 \end{matrix}\right\}$

(where p is the probability of a success in a binomial situation with 16 trials). One observation equals 10 and hence we omit this from the sample. There are 16 other observations or we have a random sample of size 16. If the numbers greater than 10 are denoted by a plus and the numbers less than 10 are denoted by a minus, we have the following data.

10.1,	10.3,	10.5,	10.1,	10.0.	9.7,	9.8,	10.2,	9.9,	10.2,	10.2,	10.4,	10.1,
+	+	+	+	.	−	−	+	−	+	+	+	+
9.9,	9.8,	10.1	10.3.									
−	−	+	+									

x = the number of plus signs = the number of successes = 11. For a two tail binomial test at the 5% level with total number of trials equal to 16, $P\{x\geqslant 12\}\leqslant 0.025$ and $P\{x\leqslant 3\}\leqslant 0.025$.

But the observed number of successes equals 11 and this does not lie in the critical region and hence we will accept the hypothesis at the 5% level.

Comments. When the sample size is large we can use a normal approximation. (See section 10.6). The same technique

can be used for testing equality of two parameters or equality of two populations if we have paired samples. If we have n paired observations where the first observations come from one population and the second observations come from a second population then a plus sign can be assigned to the pairs for which the first observation is greater than the second one and a minus sign is assigned to the pairs for which the second observation is greater than the first or *vice versa*. Now the problem reduces to the one similar to the single sample problem. Here also the pairs for which there are ties will be omitted. The assumption of symmetry in the population can be avoided if we take θ_0 as the median in the population designated by the stochastic variable X so that $P\{x>\theta_0\}=P\{x<\theta_0\}$.

12.82. **The Rank Tests.** This test is based on the rank sums and can be conveniently used for testing the equality of populations or in other words for testing whether two samples have come from identical populations. The two sample sign test could be applied if we had the two samples of the same size. This test can be applied even if the sample sizes are different. Consider two random samples of sizes n_1 and n_2. We want to test whether the samples have come from identical populations. We pool the sample observations and rank them according to the order of magnitude. If some observations have the same magnitude distribute the mean ranks among the observations. For example if there are 3 smallest observations assign the rank $(1+2+3)/3=2$ to each of them. The next observation has rank 4 etc. One of the tests known as the Mann-Whitney U test is based on the statistic U, where,

$$U=n_1n_2+n_1(n_1+1)/2-R_1$$

where n_1 and n_2 are the sample sizes and R_1 is the sum of the ranks occupied by the first sample of size n_1. It can be shown that the mean and variance of U are,

$$E(U)=n_1n_2/2$$

and $$\sigma_u^2=n_1n_2(n_1+n_2+1)/12$$

and further the standardized statistic

$$T=[U-E(U)]/\sigma_u$$

is approximately normally distributed when n_1 and n_2 are large. A good approximation is obtained when n_1 and n_2 are greater than 8. Exact tables of U are given in D.B. Owen, Handbook of Statistical Tables, Addison Wesley, 1962 (See the bibliography at the end of this chapter).

Ex. 12.82.1. *The following data give the increase in weights of samples of 10 and 9 experimental animals who are given two diets A and B. Test the hypothesis, at the 5% level, that the diets are equally effective by testing the hypothesis that the samples have come from identical populations.*

Diet A	*12*	*15*	*14*	*13*	*12*	*11*	*10*	*16*	*17*	*18*
Diet B	*10*	*9*	*8*	*14*	*19*	*20*	*21*	*22*	*22*	

Sol. We will pool the samples and will arrange the numbers according to the order of their magnitudes. In order to distinguish the numbers we will use a subscript A for the numbers from diet A (first sample).

Pooled sample	8	9	10	10_A	11_A	12_A	12_A	13_A	14_A	14	15_A
Ranks	1	2	3.5	3.5	5	6.5	6.5	8	9.5	9.5	11

Pooled sample	16_A	17_A	18_A	19	20	21	22	22
Ranks	12	13	14	15	16	17	18.5	18.5

Here for example there are two numbers equal to 10 which are supposed to occupy the ranks 3 and 4 and hence they are given the ranks $(3+4)/2=3\ 5$ each. The total number of ranks occupied by the first sample

$$=3.5+5+6.5+6.5+8+9.5+11+12+13+14=89=R_1$$

Therefore an observed U

$$=n_1n_2+n_1(n_1+1)/2-R_1=(10)(9)+(10)(11)/2-89=56.$$

$$E(U)=n_1n_2/2=(10)(9)/2=45$$

and $$\sigma_u^2=n_1n_2(n_1+n_2+1)/12$$

$$=(10)(9)(10+9+1)/12=150$$

An observed value of the standardized U

$$=t=[u-E(U)]/\sigma_u$$

$$=(56-45)/\sqrt{150}<1.96$$

Since T has an approximate normal distribution,

$$P\{t \geqslant 1.96\}=0.025$$

approximately. Hence the observed T does not fall in the critical region and hence the hypothesis that the populations are identical cannot be rejected from the evidence of these two samples.

Comments. This U test can be modified to test the hypothesis that two populations are identical, against the alternative

that the population variances are different. In testing the hypothesis that k independent samples have come from identical populations, a test known as Kruskal-Wallis H-test, is usually used. Another important distribution free test is the run test. For these and related matters see the bibliography at the end of this chapter.

Exercises

12.36. Use a sign test for testing the hypothesis, at the 5% level, that the mean yield of a hybrid corn is 50 units, by using the following data which give the yield of corn in 18 test plots. 48, 48.5, 49, 49, 50, 52, 57, 54, 53.5, 52, 49.5, 53, 52.5, 51, 47, 52, 51.5, 53.

12.37. A test conducted on 41 experimental animals to study the interval of time from the time they are subjected to a particular experimental condition till they die, yields the following time intervals. 1.92, 1.93, 1.92, 1.98, 2.00, 2.00, 2.20, 2.10, 2.20, 2.15, 2.17, 2.18, 2.20, 2.21. 2.30, 2.18, 2.10, 1.99, 1.98, 2.30, 2.22, 2.35, 2.25, 2.17, 2.12, 2.18, 2.19, 1.97, 1.98, 1.96, 2.10, 2.12, 2.15, 2.24, 2.30, 2.18, 1.94, 1.95, 1.96, 198, 1.9. Use a sign test to test the hypothesis that the expected duration $\theta=2$ against the alternative that $\theta>2$ at the 1% level.

12.38. Two types of missiles are test fired 12 times. The flight distances are given below. Pair the observations at random and use a two sample sign test at 5% level, to test the hypothesis that the two types of missiles are equally good as far as the expected flight distances are concerned.

A 2000, 2050, 2045, 2100, 2075, 2070, 2080, 2050, 2030, 2040, 2075, 2055.

B 1920, 1980, 2100, 2055, 2040, 2025, 2015. 1990, 2045, 2060, 2025, 2035

12.39. Out of 40 tea tasters who have tasted two different brands of tea, 25 of them preferred brand A, 12 of them preferred brand B and the rest could not prefer one to the other. Use a 5% sign test to test the hypothesis that brand A is better than brand B.

12.40. Kruskal-Wallis H test. This test is similar to the sign test and is used for testing whether k independent samples have come from k identical populations. The samples are pooled and ranked. The statistic used is,

$$H=\frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}-3(n+1) \text{ where } n=\sum_{i=1}^{k} n_i,$$

n_i is the size of the ith sample and R_i is the sum of the ranks occupied by the ith sample. When $n_i>5$ for all i, H has an approximate chi-square distribution with $k-1$ degrees of freedom under the null hypothesis.

The following data give the percentage reduction in skin rash by the help of three different beauty treatments conducted on three random samples of girls of a particular category. Test the hypothesis that the three treatments are equally effective, by using a H test, at 1% level.

Treatment A.	20	15	18	19	20	22	16	18	20	17
Treatment B.	18	16	15	22	21	20	18			
Treatment C.	21	20	18	17	18	22				

12.41. The Run Test. This is a convenient test for testing the randomness of a sample after having obtained the sample. The test is based

on the runs in the sample. If there is a sequence of two letters A and Bnt. on a succession of identical symbols is called a run. For example in the seque he A B B B A A B A there are 5 runs. If in a sequence there are n_1 symbolses one type and n_2 symbols of the second type (the order in which the syrabolfc occur does not matter) then the statistic = the standardized run

$$=T=[R-E(R)]/\sigma_R$$

can be shown to have approximately a standard normal distribution wher E(R) and σ_R can be shown to be,

$$E(R)=2n_1n_2/(n_1+n_2)+1 \text{ and}$$

$$\sigma^2_R=2n_1n_2(2n_1n_2-n_1-n_2)/(n_1+n_2)^2(n_1+n_2-1).$$

Hence we can test the randomness of the sample by using this norma approximation. A good approximation is obtained when n_1 and n_2 arel greater than 10. Instead of a qualitative data as described above, if we haves a numerical data we can easily put the data into a sequence of two letter- A and B where A denotes the numbers greater than the median and B de notes the number less than the median.

A machine produces a particular article which can be classified as defective D or non-defective G. One item each on every half hour is tested for quality. The data is given below. Test at 1% level, whether the sample can be considered to be random as far as the quality specifications of the article are concerned. G G G D G G D D G D G G G D D D G D G D G G G G D D D G G D G G G D D D G D G D G G G G G D D G.

BIBLIOGRAPHY AND REFERENCES

1. F.S. **Acton,** Analysis of Straight-line Data. Wiley, New York, 1959.

2. H.D. **Brunk,** An Introduction to Mathematical Statistics, Blaisdell Publishing Co. New York, 1965.

3. H. **Chernoff** and L.E. **Moses,** Elementary Decision Theory. Wiley, New York, 1959.

4. W.G. **Cochran** and G.M. **Cox,** Experimental Designs. Wiley, New York, 1957.

5. H. **Cramer,** Mathematical Methods of Statistics, Princeton, University Press, 1946.

6. D.K. **Faddeev** and V.N. **Faddeeva,** Computational Methods of Linear Algebra. Translated from Russian, Freeman & Co. San Francisco, 1963.

7. F.A. **Graybill,** Linear Statistical Models. McGraw-Hill, New York, 1961.

8 I. **Guttman,** and S.S. **Wilks,** Introductory Engineering Statistics, Wiley, New York, 1965.

9. P.G. **Hoel,** Introduction to Mathematical Statistics, Wiley, New York, 1962.

10. O. **Kempthorne,** The Design and Analysis of Experiments. Wiley, New York, 1952.

11. M.G. **Kendall** and A. **Stuart,** The Advanced Theory of Statistics. Vol. 2. Charles Griffin & Co. London, 1961.

12. M.G **Kendall,** Regression Structure and Functional Relationship. Part I. Biometrika, Vol. 38, 1951, pp. 11-25.

13. C.C. **Li,** Introduction to Experimental Statistics. McGraw-Hill, New York, 1964.

14. D.V. **Dindley,** Regression Lines and Linear Functional relationships. Supp. J. Royal Stat. Soc. Vol. 9. 1947.

15. A.M. **Mathai,** An approximate method of analysis for a two-way layout. Biometrics. Vol. 21. 1965 pp. 376—385.

16. A.M. **Mathai,** Measures of Dispersion. Unpublished manuscript, 1965.

17. A.M. **Mathai,** Some characterizations of one-parameter family of distributions. Canadian Math. Bull. Vol. 9, 1966 pp. 95—102.

18. A.M. **Mathai,** Sampling Distributions under missing values. Trabajos de Estadistica, vol. 17, 1966, pp. 59—83.

19. A.M. **Mathai** and R.K. **Saxena,** On a generalized hypergeometric distribution, Metrika, Vol. 11, 1967, pp. 127—132.

20. A.M. **Mathai** and R.K. **Saxena,** Distributions having the gamma properties. Metrika, Vol. 11, 1967 (to appear)

21. A.M. **Mathai** and R.K. **Saxena,** A short table of generalized hypergeometric distribution. Metrika. Vol. 12, 1968 (to appear).

22. A.M. **Mathai,** Pseudo-inverses in normal equations. *Estadistica*. 1960. No. 92.

23. A.M. **Mathai**, Dispersion theory. *Estadistica* 1967 (to appear).

24. A.M. **Mathai**, On a problem of regression. *S. Afr. Statist.* J. Vol. 1, 1967 (to appear).

25. A.M. **Mathai**, Dispersion and information. *Metron*. 1967 (to appear).

26. A.M. **Mathai**, On the structural properties of conditional distributions. Canad. Math. Bull. Vol. 10, 1967 (to appear).

27. G.H. **Moore,** and W.A. **Wallis,** Time series significance tests based on signs of differences. J. Amer. Stat. Assoc. Vol. 38, 1943, pp. 153-164.

28. K.R. **Nair** and K.S. **Banerjee,** A note on fitting straight lines if both variables are subject to errors. Sankhya, Vol. 6, 1942-43, pp. 331.

29. D.B. **Owen,** Handbook of Statistical Tables Addison Wesley, Reading, Mass., 1962.

30. H. **Scheffe,** The Analysis of Variance. Wiley, New York, 1959.

31. F. **Wilcoxin,** Individual comparisons by ranking methods. Biometrics. Bull. Vol. 1, 1945, pp. 80-83.

32. R.L. **Wine,** Statistics for Scientists and Engineers. Prentice-Hall, Englewood Cliffs, N.J. 1964.

TABLE 1*

BINOMIAL COEFFICIENTS

$$\text{Entry}: \binom{n}{x} = \binom{n}{n-x} = \frac{n!}{x!\,(n-x)!}$$

$n \backslash x$	1	2	3	4	5	6	7	8	9	10	11
5	1	5	10								
6	1	6	15	20							
7	1	7	21	35							
8	1	8	28	56	70						
9	1	9	36	84	126						
10	1	10	45	120	210	252					
11	1	11	55	165	330	462					
12	1	12	66	220	495	792	924				
13	1	13	78	286	715	1287	1716				
14	1	14	91	364	1001	2002	3003	3432			
15	1	15	105	455	1365	3003	5005	6435			
16	1	16	120	560	1820	4368	8008	11440	12870		
17	1	17	136	680	2380	6188	12376	19448	24310		

n \ x	1	2	3	4	5	6	7	8	9	10	11	12
18	1	18	153	816	3060	8568	18564	31824	43758	48620		
19	1	19	171	969	3876	11628	27132	50388	75582	92378		
20	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756	
21	1	21	210	1330	5985	20349	54264	116280	203490	293930	352716	
22	1	22	231	1540	7315	26334	74613	170544	319770	497420	646646	705432
23	1	23	253	1771	8855	33649	100947	245157	490314	817190	1144066	1352078
24	1	24	276	2024	10626	42504	134596	346104	735471	1307504	1961256	2496144
25	1	25	300	2300	12650	53130	177100	480700	1081575	2042975	3268760	4457400
26	1	26	325	2600	14950	65780	230230	657800	1562275	3124550	5311735	7726160
27	1	27	351	2925	17550	80730	296010	888030	2220075	4686825	8436285	13037895
28	1	28	378	3276	20475	98280	376740	1184040	3108105	6906900	13123110	21474180
29	1	29	406	3654	23751	118755	475020	1560780	4292145	10015005	20030010	34597290
30	1	30	435	4060	27405	142506	593775	2035800	5852925	14307150	30045015	54627300

$n \backslash x$	13	14	15	16
24	2704156			
25	5200300			
26	9657700	10400600		
27	17383860	20058300		
28	30421755	37442160	40116600	
29	51895935	67863915	77558760	
30	86493225	119759850	145422675	155117520

*The author would like to thank Mr. Morty Yalovsky for computing this table, and the McGill University, Department of Computer Science for computer facilities.

TABLE 2*

CUMULATIVE BINOMIAL PROBABILITIES

Entry : $\sum_{r=0}^{x} \binom{n}{r} p^r (1-p)^{n-r}$

n	x	$p=$ .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
1	0	.9500	.9000	.8500	.8000	.7500	.7000	.6500	.6000	.5500	.5000
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0	.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500
	1	.9975	.9900	.9775	.9600	.9375	.9100	.8775	.8400	.7975	.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250
	1	.9927	.9720	.9392	.8960	.8437	.7840	.7182	.6480	.5747	.5000
	2	.9999	.9990	.9966	.9920	.9844	.9730	.9571	.9360	.9089	.8750
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	.8145	.6561	.5220	.4096	.3164	.2401	17.85	.1296	.0915	.0625
	1	.9860	.9477	.8905	.8192	.7383	.6517	.5630	.4752	.3910	.3125
	2	.9995	.9963	.9880	.9728	.9492	.9163	.8735	.8208	.7585	.6875
	3	1.0000	.9999	.9995	.9984	.9961	.9919	.9850	.9744	.9590	.9375
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

5	0	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0313
	1	.9774	.9185	.8352	.7373	.6328	.5282	.4284	.3370	.2562	.1875
	2	.9988	.9914	.9734	.9421	.8965	.8369	.7648	.6826	.5931	.5000
	3	1.0000	.9995	.9978	.9933	.9844	.9692	.9460	.9130	.8688	.8125
	4	1.0000	1.0000	.9999	.9997	.9990	.9976	.9947	.9898	.9815	.9688
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	.7351	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277	.0156
	1	.9672	.8857	.7765	.6554	.5339	.4202	.3191	.2333	.1636	.1094
	2	.9978	.9841	.9527	.9011	.8306	.7443	.6471	.5443	.4415	.3438
	3	.9999	.9987	.9941	.9830	.9624	.9295	.8826	.8208	.7447	.6562
	4	1.0000	.9999	.9996	.9984	.9954	.9891	.9777	.9590	.9308	.8906
	5	1.0000	1.0000	1.0000	.9999	.9998	.9993	.9982	.9959	.9917	.9844
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0	.6983	.4783	.3206	.2097	.1335	.0824	.0490	.0280	.0152	.0078
	1	.9556	.8503	.7166	.5767	.4449	.3294	.2338	.1586	.1024	.0625
	2	.9962	.9743	.9262	.8520	.7564	.6471	.5323	.4199	.3164	.2266
	3	.9998	.9973	.9879	.9667	.9294	.8740	.8002	.7102	.6083	.5000
	4	1.0000	.9998	.9988	.9953	.9871	.9712	.9444	.9037	.8471	.7734
	5	1.0000	1.0000	.9999	.9996	.9987	.9962	.9910	.9812	.9643	.9375
	6	1.0000	1.0000	1.0000	1.0000	.9999	.9998	.9994	.9984	.9963	.9922
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	0	.6634	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084	.0039
	1	.9428	.8131	.6572	.5033	.3671	.2553	.1691	.1064	.0632	.0352
	2	.9942	.9619	.8948	.7969	.6785	.5518	.4278	.3154	.2201	.1445
	3	.9996	.9950	.9786	.9437	.8862	.8059	.7064	.5941	.4770	.3633
	4	1.0000	.9996	.9971	.9896	.9727	.9420	.8939	.8263	.7396	.6367
	5	1.0000	1.0000	.9998	.9988	.9958	.9887	.9747	.9502	.9115	.8555

n	x	$p=$ .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
	6	1.0000	1.0000	1.0000	.9999	.9996	.9987	.9964	.9915	.9819	.9648
	7	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9998	.9993	.9983	.9961
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046	.0020
	1	.9288	.7748	.5995	.4362	.3003	.1960	.1211	.0705	.0385	.0195
	2	.9916	.9470	.8591	.7382	.6007	.4628	.3373	.2318	.1495	.0898
	3	.9994	.9917	.9661	.9144	.8343	.7297	.6089	.4826	.3614	.2539
	4	1.0000	.9991	.9944	.9804	.9511	.9012	.8283	.7334	.6214	.5000
	5	1.0000	.9999	.9994	.9969	.9900	.9747	.9464	.9006	.8342	.7461
	6	1.0000	1.0000	1.0000	.9997	.9987	.9957	.9888	.9750	.9502	.9102
	7	1.0000	1.0000	1.0000	1.0000	.9999	.9996	.9986	.9962	.9909	.9805
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9992	.9980
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010
	1	.9139	.7361	.5443	.3758	.2440	.1493	.0860	.0464	.0233	.0107
	2	9885	.9298	.8202	.6778	.5256	.3828	.2616	.1673	.0996	.0547
	3	.9990	.9872	.9500	.8791	.7759	.6496	.5138	.3823	.2660	.1719
	4	.9999	.9984	.9901	.9672	.9219	.8497	.7515	.6331	.5044	.3770
	5	1.0000	.9999	.9986	.9936	.9803	.9527	.9051	.8338	.7384	.6230
	6	1.0000	1.0000	.9999	.9991	.9965	.9894	.9740	.9452	.8980	.8281
	7	1.0000	1.0000	1.0000	.9999	.9996	.9984	.9952	.9877	.9726	.9453
	8	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9995	.9983	.9955	.9893
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9990
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

11	0	.5688	.3138	.1673	.0859	.0422	.0198	.0088	.0036	.0014	.0005
	1	.8981	.6974	.4922	.3221	.1971	.1130	.0606	.0302	.0139	.0059
	2	.9848	.9104	.7788	.6174	.4552	.3127	.2001	.1189	.0652	.0327
	3	.9984	.9815	.9306	.8389	.7133	.5696	.4256	.2963	.1911	.1133
	4	.9999	.9972	.9841	.9496	.8854	.7897	.6683	.5328	.3971	.2744
	5	1.0000	.9997	.9973	.9883	.9657	.9218	.8513	.7535	.6331	.5000
	6	1.0000	1.0000	.9997	.9980	.9924	.9784	.9499	.9006	.8262	.7256
	7	1.0000	1.0000	1.0000	.9998	.9988	.9957	.9878	.9707	.9390	.8867
	8	1.0000	1.0000	1.0000	1.0000	.9999	.9994	.9980	.9941	.9852	.9673
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9998	.9993	.9978	.9941
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9998	.9995
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
12	0	.5404	.2824	.1422	.0687	.0317	.0138	.0057	.0022	.0008	.0002
	1	.8816	.6590	.4435	.2749	.1584	.0850	.0424	.0196	.0083	.0032
	2	.9804	.8891	.7358	.5583	.3907	.2528	.1513	.0834	.0421	.0193
	3	.9978	.9744	.9078	.7946	.6488	.4925	.3467	.2253	.1345	.0730
	4	.9998	.9957	.9761	.9274	.8424	.7237	.5833	.4382	.3044	.1938
	5	1.0000	.9995	.9954	.9806	.9456	.8822	.7873	.6652	.5269	.3872
	6	1.0000	.9999	.9993	.9961	.9857	.9614	.9154	.8418	.7393	.6128
	7	1.0000	1.0000	.9999	.9994	.9972	.9905	.9745	.9427	.8883	.8062
	8	1.0000	1.0000	1.0000	.9999	.9996	.9983	.9944	.9847	.9644	.9270
	9	1.0000	1.0000	1.0000	1.0000	1.0000	.9998	.9992	.9972	.9921	.9807
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9989	.9968
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9998
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n	n	p= .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
13	0	.5133	.2542	.1209	.0550	.0238	.0097	.0037	.0013	.0004	.0001
	1	.8646	.6213	.3983	.2336	.1267	.0637	.0296	.0126	.0049	.0017
	2	.9755	.8661	.6920	.5017	.3326	.2025	.1132	.0579	.0269	.0112
	3	.9969	.9658	.8820	.7473	.5843	.4206	.2783	.1686	.0929	.0461
	4	.9997	.9935	.9658	.9009	.7940	.6543	.5005	.3530	.2279	.1334
	5	1.0000	.9991	.9925	.9700	.9198	.8346	.7159	.5744	.4268	.2905
	6	1.0000	.9999	.9987	.9930	.9757	.9376	.8705	.7712	.6437	.5000
	7	1.0000	1.0000	.9998	.9988	.9944	.9818	.9538	.9023	.8212	.7095
	8	1.0000	1.0000	1.0000	.9998	.9990	.9960	.9874	.9679	.9302	.8666
	9	1.0000	1.0000	1.0000	1.0000	.9999	.9993	.9975	.9922	.9797	.9539
	10	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9987	.9959	.9888
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9995	.9983
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	0	.4877	.2288	.1028	.0440	.0178	.0068	.0024	.0008	.0002	.0001
	1	.8470	.5846	.3567	.1979	.1010	.0475	.0205	.0081	.0029	.0009
	2	.9699	.8416	.6479	.4481	.2811	.1608	.0839	.0398	.0170	.0065
	3	.9958	.9559	.8535	.6982	.5213	.3552	.2205	.1243	.0632	.0287
	4	.9996	.9908	.9533	.8702	.7415	.5842	.4227	.2793	.1672	.0898
	5	1.0000	.9985	.9885	.9561	.8883	.7805	.6405	.4859	.3373	.2120
	6	1.0000	.9998	.9978	.9884	.9617	.9064	.8164	.6925	.5461	.3953
	7	1.0000	1.0000	.9997	.9976	.9897	.9685	.9247	.8499	.7414	.6047
	8	1.0000	1.0000	1.0000	.9996	.9978	.9917	.9757	.9417	.8811	.7880
	9	1.0000	1.0000	1.0000	1.0000	.9997	.9983	.9940	.9825	.9574	.9102
	10	1.0000	1.0000	1.0000	1.0000	1.0000	.9998	.9989	.9961	.9886	.9713
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9994	.9978	.9935
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9991
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

15	0	.4633	.2059	.0874	.0352	.0134	.0047	.0016	.0005	.0001	.0000
	1	.8290	.5490	.3186	.1671	.0802	.0353	.0142	.0052	.0017	.0005
	2	.9638	.8159	.6042	.3980	.2361	.1268	.0617	.0271	.0107	.0037
	3	.9945	.9444	.8227	.6482	.4613	.2969	.1727	.0905	.0424	.0176
	4	.9994	.9873	.9383	.8358	.6865	.5155	.3519	.2173	.1204	.0592
	5	.9999	.9978	.9832	.9389	.8516	.7216	.5643	.4032	.2608	.1509
	6	1.0000	.9997	.9964	.9819	.9434	.8689	.7548	.6098	.4522	.3036
	7	1.0000	1.0000	.9994	.9958	.9827	.9500	.8868	.7869	.6535	.5000
	8	1.0000	1.0000	.9999	.9992	.9958	.9848	.9578	.9050	.8182	.6964
	9	1.0000	1.0000	1.0000	.9999	.9992	.9963	.9876	.9662	.9231	.8491
	10	1.0000	1.0000	1.0000	1.0000	.9999	.9993	.9972	.9907	.9745	.9408
	11	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9995	.9981	.9937	.9824
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9989	.9963
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
16	0	.4401	.1853	.0743	.0281	.0100	.0033	.0010	.0003	.0001	.0000
	1	.8108	.5147	.2839	.1407	.0635	.0261	.0098	.0033	.0010	.0003
	2	.9571	.7892	.5614	.3518	.1971	.0994	.0451	.0183	.0066	.0021
	3	.9930	.9316	.7899	.5981	.4050	.2459	.1339	.0651	.0281	.0106
	4	.9991	.9830	.9209	.7982	.6302	.4499	.2892	.1666	.0853	.0384
	5	.9999	.9967	.9765	.9183	.8103	.6598	.4900	.3288	.1976	.1051
	6	1.0000	.9995	.9944	.9733	.9204	.8247	.6881	.5272	.3660	.2272
	7	1.0000	.9999	.9989	.9930	.9729	.9256	.8406	.7161	.5629	.4018
	8	1.0000	1.0000	.9998	.9985	.9925	.9743	.9329	.8577	.7441	.5982
	9	1.0000	1.0000	1.0000	.9998	.9984	.9929	.9771	.9417	.8759	.7728
	10	1.0000	1.0000	1.0000	1.0000	.9997	.9984	.9938	.9809	.9514	.8949
	11	1.0000	1.0000	1.0000	1.0000	1.0000	.9997	.9987	.9951	.9851	.9616
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9998	.9991	.9965	.9894
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9994	.9979
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n	x	p= .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
17	0	.4181	.1668	.0631	.0225	.0075	.0023	.0007	.0002	.0000	.0000
	1	.7922	.4818	.2525	.1182	.0501	.0193	.0067	.0021	.0006	.0001
	2	.9497	.7618	.5198	.3096	.1637	.0774	.0327	.0123	.0041	.0012
	3	.9912	.9174	.7556	.5489	.3530	.2019	.1028	.0464	.0184	.0064
	4	.9988	.9779	.9013	.7582	.5739	.3887	.2348	.1260	.0596	.0245
	5	.9999	.9953	.9681	.8943	.7653	.5968	.4197	.2639	.1471	.0717
	6	1.0000	.9992	.9917	.9623	.8929	.7752	.6188	.4478	.2902	.1662
	7	1.0000	.9999	.9983	.9891	.9598	.8954	.7872	.6405	.4743	.3145
	8	1.0000	1.0000	.9997	.9974	.9876	.9597	.9006	.8011	.6626	.5000
	9	1.0000	1.0000	1.0000	.9995	.9969	.9873	.9617	.9081	.8166	.6855
	10	1.0000	1.0000	1.0000	.9999	.9994	.9968	.9880	.9652	.9174	.8338
	11	1.0000	1.0000	1.0000	1.0000	.9999	.9993	.9970	.9894	.9699	.9283
	12	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9994	.9975	.9914	.9755
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9995	.9981	.9936
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9988
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	0	.3972	.1501	.0536	.0180	.0056	.0016	.0004	.0001	.0000	.0000
	1	.7735	.4503	.2241	.0991	.0395	.0142	.0046	.0013	.0003	.0001
	2	.9419	.7338	.4797	.2713	.1353	.0600	.0236	.0082	.0025	.0007
	3	.9891	.9018	.7202	.5010	.3057	.1646	.0783	.0328	.0120	.0038
	4	.9985	.9718	.8794	.7164	.5187	.3327	.1886	.0942	.0411	.0154
	5	.9998	.9936	.9581	.8671	.7175	.5344	.3550	.2088	.1077	.0481
	6	1.0000	.9988	.9882	.9487	.8610	.7217	.5491	.3743	.2258	.1189
	7	1.0000	.9998	.9973	.9837	.9431	.8593	.7283	.5634	.3915	.2403
	8	1.0000	1.0000	.9995	.9957	.9807	.9404	.8609	.7368	.5778	.4073
	9	1.0000	1.0000	.9999	.9991	.9946	.9790	.9403	.8653	.7473	.5927
	10	1.0000	1.0000	1.0000	.9998	.9988	.9939	.9788	.9424	.8720	.7597

	11	1.0000	1.0000	1.0000	1.0000	.9998	.9986	.9938	.9797	.9463	.8811
	12	1.0000	1.0000	1.0000	1.0000	1.0000	.9997	.9986	.9942	.9817	.9519
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9997	.9987	.9951	.9846
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9998	.9990	.9962
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9993
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
19	0	.3774	.1351	.0456	.0144	.0042	.0011	.0003	.0001	.0000	.0000
	1	.7547	.4203	.1985	.0829	.0310	.0104	.0031	.0008	.0002	.0000
	2	.9335	.7054	.4413	.2369	.1113	.0462	.0170	.0055	.0015	.0004
	3	.9868	.8850	.6841	.4551	.2631	.1332	.0591	.0230	.0077	.0022
	4	.9980	.9648	.8556	.6733	.4654	.2822	.1500	.0696	.0280	.0096
	5	.9998	.9914	.9463	.8369	.6678	.4739	.2968	.1629	.0777	.0318
	6	1.0000	.9983	.9837	.9324	.8251	.6655	.4812	.3081	.1727	.0835
	7	1.0000	.9997	.9959	.9767	.9225	.8180	.6656	.4878	.3169	.1796
	8	1.0000	1.0000	.9992	.9933	.9713	.9161	.8145	.6675	.4940	.3238
	9	1.0000	1.0000	.9999	.9984	.9911	.9674	.9125	.8139	.6710	.5000
	10	1.0000	1.0000	1.0000	.9997	.9977	.9895	.9653	.9115	.8159	.6762
	11	1.0000	1.0000	1.0000	1.0000	.9995	.9972	.9886	.9648	.9129	.8204
	12	1.0000	1.0000	1.0000	1.0000	.9999	.9994	.9969	.9884	.9658	.9165
	13	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9993	.9969	.9891	.9682
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9994	.9972	.9904
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9995	.9978
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9996
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n	x	p= .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
	0	.3585	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000	.0000
	1	.7358	.3917	.1756	.0692	.0243	.0076	.0021	.0005	.0001	.0000
	2	.9245	.6769	.4049	.2061	.0913	.0355	.0121	.0036	.0009	.0002
	3	.9841	.8670	.6477	.4114	.2252	.1071	.0444	.0160	.0049	.0013
	4	.9974	.9568	.8298	.6296	.4148	.2375	.1182	.0510	.0189	.0059
	5	.9997	.9887	.9327	.8042	.6172	.4164	.2454	.1256	.0553	.0207
	6	1.0000	.9976	.9781	.9133	.7858	.6080	.4166	.2500	.1299	.0577
	7	1.0000	.9996	.9941	.9679	.8982	.7723	.6010	.4159	.2520	.1316
	8	1.0000	.9999	.9987	.9900	.9591	.8867	.7624	.5956	.4143	.2517
	9	1.0000	1.0000	.9998	.9974	.9861	.9520	.8782	.7553	.5914	.4119
	10	1.0000	1.0000	1.0000	.9994	.9961	.9829	.9468	.8725	.7507	.5881
	11	1.0000	1.0000	1.0000	.9999	.9991	.9949	.9804	.9435	.8692	.7483
	12	1.0000	1.0000	1.0000	1.0000	.9998	.9987	.9940	.9790	.9420	.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	.9997	.9985	.9935	.9786	.9423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9997	.9984	.9936	.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9985	.9941
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9997	.9987
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

*The author would like to thank Mr. Morty Yalovsky for computing this table, and the McGill University, Department of Computer Science for computer facilities.

... facilities. ... for computing this table, and the McGill University, Department of

TABLE 3*

CUMULATIVE POISSON PROBABILITIES

$$\text{Entry}: \sum_{r=0}^{x} \frac{\lambda^r}{r!} e^{-\lambda}$$

x \ λ	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
0	.9048	.8187	.7408	.6703	.6065	.5488	.4966	.4493	.4066	.3679
1	.9953	.9825	.9631	.9384	.9098	.8781	.8442	.8088	.7725	.7358
2	.9998	.9989	.9964	.9921	.9856	.9769	.9659	.9526	.9371	.9197
3	1.0000	.9999	.9997	.9992	.9982	.9966	.9942	.9909	.9865	.9810
4	1.0000	1.0000	1.0000	.9999	.9998	.9996	.9992	.9986	.9977	.9963
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9998	.9997	.9994
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x \ λ	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	.3329	.3012	.2725	.2466	.2231	.2019	.1827	.1653	.1496	.1353
1	.6990	.6626	.6268	.5918	.5578	.5249	.4932	.4628	.4337	.4060
2	.9004	.8795	.8571	.8335	.8088	.7834	.7572	.7306	.7037	.6767
3	.9743	.9662	.9569	.9463	.9344	.9212	.9068	.8913	.8747	.8571
4	.9946	.9923	.9893	.9857	.9814	.9763	.9704	.9636	.9559	.9473
5	.9990	.9985	.9978	.9968	.9955	.9940	.9920	.9896	.9868	.9834
6	.9999	.9997	.9996	.9994	.9991	.9987	.9981	.9974	.9966	.9955
7	1.0000	1.0000	.9999	.9999	.9998	.9997	.9996	.9994	.9992	.9989
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9999	.9998	.9998
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

$x \backslash \lambda$	2.1	2.2	2.3	[2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	.1225	.1108	.1003	.0907	.0821	.0743	.0672	.0608	.0550	.0498
1	.3796	.3546	.3309	.3084	.2873	.2674	.2487	.2311	.2146	.1991
2	.6496	.6227	.5960	.5697	.5438	.5184	.4936	.4695	.4460	.4232
3	.8386	.8194	.7993	.7787	.7576	.7360	.7141	.6919	.6696	.6472
4	.9379	.9275	.9162	.9041	.8912	.8774	.8629	.8477	.8318	.8153
5	.9796	.9751	.9700	.9643	.9580	.9510	.9433	.9349	.9258	.9161
6	.9941	.9925	.9906	.9884	.9858	.9828	.9794	.9756	.9713	.9665
7	.9985	.9980	.9974	.9967	.9958	.9947	.9934	.9919	.9901	.9881
8	.9997	.9995	.9994	.9991	.9989	.9985	.9981	.9976	.9969	.9962
9	.9999	.9999	.9999	.9998	.9997	.9996	.9995	.9993	.9991	.9989
10	1.0000	1.0000	1.0000	1.0000	.9999	.9999	.9999	.9998	.9998	.9997
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9999
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

$x \backslash \lambda$	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	.0450	.0408	.0369	.0334	.0302	.0273	.0247	.0224	.0202	.0183
1	.1847	.1712	.1586	.1468	.1359	.1257	.1162	.1074	.0992	.0916
2	.4012	.3799	.3594	.3397	.3208	.3027	.2854	.2689	.2531	.2381
3	.6248	.6025	.5803	.5584	.5366	.5152	.4942	.4735	.4532	.4335
4	.7982	.7806	.7626	.7442	.7254	.7064	.6872	.6678	.6484	.6288
5	.9057	.8946	.8829	.8705	.8576	.8441	.8301	.8156	.8006	.7851
6	.9612	.9554	.9490	.9421	.9347	.9267	.9182	.9091	.8995	.8893
7	.9858	.9832	.9802	.9769	.9733	.9692	.9648	.9599	.9546	.9489
8	.9953	.9943	.9931	.9917	.9901	.9883	.9863	.9840	.9815	.9786
9	.9986	.9982	.9978	.9973	.9967	.9960	.9952	.9942	.9931	.9919
10	.9996	.9995	.9994	.9992	.9990	.9987	.9984	.9981	.9977	.9972
11	.9999	.9999	.9998	.9998	.9997	.9996	.9995	.9994	.9993	.9991
12	1.0000	1.0000	1.0000	.9999	.9999	.9999	.9999	.9998	.9998	.9997
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9999
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x \ λ	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	.0166	.0150	.0136	.0123	.0111	.0101	.0091	.0082	.0074	.0067
1	.0845	.0780	.0719	.0663	.0611	.0563	.0518	.0477	.0439	.0404
2	.2238	.2102	.1974	.1851	.1736	.1626	.1523	.1425	.1333	.1247
3	.4142	.3954	.3772	.3594	.3423	.3257	.3097	.2942	.2793	.2650
4	.6093	.5898	.5704	.5512	.5321	.5132	.4946	.4763	.4582	.4405
5	.7693	.7531	.7367	.7199	.7029	.6858	.6684	.6510	.6335	.6160
6	.8786	.8675	.8558	.8436	.8311	.8180	.8046	.7908	.7767	.7622
7	.9427	.9361	.9290	.9214	.9134	.9049	.8960	.8867	.8769	.8666
8	.9755	.9721	.9683	.9642	.9597	.9549	.9497	.9442	.9382	.9319
9	.9905	.9889	.9871	.9851	.9829	.9805	.9778	.9749	.9717	.9682
10	.9966	.9959	.9952	.9943	.9933	.9922	.9910	.9896	.9880	.9863
11	.9989	.9986	.9983	.9980	.9976	.9971	.9966	.9960	.9953	.9945
12	.9997	.9996	.9995	.9993	.9992	.9990	.9988	.9986	.9983	.9980
13	.9999	.9999	.9998	.9998	.9997	.9997	.9996	.9995	.9994	.9993
14	1.0000	1.0000	1.0000	.9999	.9999	.9999	.9999	.9999	.9998	.9998
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9999

x \ λ	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
0	.0061	.0055	.0050	.0045	.0041	.0037	.0033	.0030	.0027	.0025
1	.0372	.0342	.0314	.0289	.0266	.0244	.0224	.0206	.0189	.0174
2	.1165	.1088	.1016	.0948	.0884	.0824	.0768	.0715	.0666	.0620
3	.2513	.2381	.2254	.2133	.0217	.1906	.1800	.1700	.1604	.1512
4	.4231	.4061	.3895	.3733	.3575	.3422	.3272	.3127	.2987	.2851

x \ λ	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
5	.5984	.5809	.5635	.5461	.5289	.5119	.4950	.4783	.4619	.4457
6	.7474	.7324	.7171	.7017	.6860	.6703	.6544	.6384	.6224	.6063
7	.8560	.8449	.8335	.8217	.8095	.7970	.7841	.7710	.7576	.7440
8	.9252	.9181	.9106	.9027	.8944	.8857	.8766	.8672	.8574	.8472
9	.9644	.9603	.9559	.9512	.9462	.9409	.9352	.9292	.9228	.9161
10	.9844	.9823	.9800	.9775	.9747	.9718	.9686	.9651	.9614	.9574
11	.9937	.9927	.9916	.9904	.9890	.9875	.9859	.9841	.9821	.9799
12	.9976	.9972	.9967	.9962	.9955	.9949	.9941	.9932	.9922	.9912
13	.9992	.9990	.9988	.9986	.9983	.9980	.9977	.9973	.9969	.9964
14	.9997	.9997	.9996	.9995	.9994	.9993	.9991	.9990	.9988	.9986
15	.9999	.9999	.9999	.9998	.9998	.9998	.9997	.9996	.9996	.9995
16	1.0000	1.0000	1.0000	.9999	.9999	.9999	.9999	.9999	.9999	.9998
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999

x \ λ	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0
0	.0022	.0020	.0018	.0017	.0015	.0014	.0012	.0011	.0010	.0009
1	.0159	.0146	.0134	.0123	.0113	.0103	.0095	.0087	.0080	.0073
2	.0577	.0536	.0498	.0463	.0430	.0400	.0371	.0344	.0320	.0296
3	.1425	.1342	.1264	.1189	.1118	.1052	.0988	.0928	.0871	.0818
4	.2719	.2592	.2469	.2351	.2237	.2127	.2022	.1920	.1823	.1730
5	.4298	.4141	.3988	.3837	.3690	.3547	.3406	.3270	.3137	.3007

6	.5902	.5742	.5582	.5423	.5265	.5108	.4953	.4799	.4647	.4497
7	.7301	.7160	.7017	.6873	.6728	.6581	.6433	.6285	.6136	.5987
8	.8367	.8259	.8148	.8033	.7916	.7796	.7673	.7548	.7420	.7291
9	.9090	.9016	.8939	.8858	.8774	.8686	.8596	.8502	.8405	.8305
10	.9531	.9486	.9437	.9386	.9332	.9274	.9214	.9151	.9084	.9015
11	.9776	.9750	.9723	.9693	.9661	.9627	.9591	.9552	.9510	.9467
12	.9900	.9887	.9873	.9857	.9840	.9821	.9801	.9779	.9755	.9730
13	.9958	.9952	.9945	.9937	.9929	.9920	.9909	.9898	.9885	.9872
14	.9984	.9981	.9978	.9974	.9970	.9966	.9961	.9956	.9950	.9943
15	.9994	.9993	.9992	.9990	.9988	.9986	.9984	.9982	.9979	.9976
16	.9998	.9997	.9997	.9996	.9996	.9995	.9994	.9993	.9992	.9990
17	.9999	.9999	.9999	.9999	.9998	.9998	.9998	.9997	.9997	.9996
18	1.0000	1.0000	1.0000	1.0000	.9999	.9999	.9999	.9999	.9999	.9999
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x \ λ	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
0	.0008	.0007	.0007	.0006	.0006	.0005	.0005	.0004	.0004	.0003
1	.0067	.0061	.0056	.0051	.0047	.0043	.0039	.0036	.0033	.0030
2	.0275	.0255	.0236	.0219	.0203	.0188	.0174	.0161	.0149	.0138
3	.0767	.0719	.0674	.0632	.0591	.0554	.0518	.0485	.0453	.0424
4	.1641	.1555	.1473	.1395	.1321	.1249	.1181	.1117	.1055	.0996
5	.2881	.2759	.2640	.2526	.2414	.2307	.2203	.2103	.2006	.1912
6	.4349	.4204	.4060	.3920	.3782	.3646	.3514	.3384	.3257	.3134
7	.5838	.5689	.5541	.5393	.5246	.5100	.4956	.4812	.4670	.4530
8	.7160	.7027	.6892	.6757	.6620	.6482	.6343	.6204	.6065	.5925
9	.8202	.8096	.7988	.7877	.7764	.7649	.7531	.7411	.7290	.7166
10	.8942	.8867	.8788	.8707	.8622	.8535	.8445	.8352	.8257	.8159
11	.9420	.9371	.9319	.9265	.9208	.9148	.9085	.9020	.8952	.8881
12	.9703	.9673	.9642	.9609	.9573	.9536	.9496	.9454	.9409	.9362
13	.9857	.9841	.9824	.9805	.9784	.9762	.9739	.9714	.9687	.9658
14	.9935	.9927	.9918	.9908	.9897	.9886	.9873	.9859	.9844	.9827
15	.9972	.9969	.9964	.9959	.9954	.9948	.9941	.9934	.9926	.9918

x \ λ	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
16	.9989	.9987	.9985	.9983	.9980	.9978	.9974	.9971	.9967	.9963
17	.9996	.9995	.9994	.9993	.9992	.9991	.9989	.9988	.9986	.9984
18	.9998	.9998	.9998	.9997	.9997	.9996	.9996	.9995	.9994	.9993
19	.9999	.9999	.9999	.9999	.9999	.9999	.9998	.9998	.9998	.9997
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9999	.9999	.9999
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x \ λ	8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8	8.9	9.0
0	.0003	.0003	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001
1	.0028	.0025	.0023	.0021	.0019	.0018	.0016	.0015	.0014	.0012
2	.0127	.0118	.0109	.0100	.0093	.0086	.0079	.0073	.0068	.0062
3	.0396	.0370	.0346	.0323	.0301	.0281	.0262	.0244	.0228	.0212
4	.0940	.0887	.0837	.0789	.0744	.0701	.0660	.0621	.0584	.0550
5	.1822	.1736	.1653	.1573	.1496	.1422	.1352	.1284	.1219	.1157
6	.3013	.2896	.2781	.2670	.2562	.2457	.2355	.2256	.2160	.2068
7	.4391	.4254	.4119	.3987	.3856	.3728	.3602	.3478	.3357	.3239
8	.5786	.5647	.5507	.5369	.5231	.5094	.4958	.4823	.4689	.4557
9	.7041	.6915	.6788	.6659	.6530	.6400	.6269	.6137	.6006	.5874
10	.8058	.7955	.7850	.7743	.7634	.7522	.7409	.7294	.7178	.7060
11	.8807	.8731	.8652	.8571	.8487	.8400	.8311	.8220	.8126	.8030
12	.9313	.9261	.9207	.9150	.9091	.9029	.8965	.8898	.8829	.8758
13	.9628	.9595	.9561	.9524	.9486	.9445	.9403	.9358	.9311	.9261
14	.9810	.9791	.9771	.9749	.9726	.9701	.9675	.9647	.9617	.9585
15	.9908	.9898	.9887	.9875	.9862	.9848	.9832	.9816	.9798	.9780
16	.9958	.9953	.9947	.9941	.9934	.9926	.9918	.9909	.9899	.9889
17	.9982	.9979	.9977	.9973	.9970	.9966	.9962	.9957	.9952	.9947

18	.9992	.9991	9990	.9989	.9987	.9985	.9983	.9981	.9978	.9976
19	.9997	.9997	.9996	.9995	.9995	.9994	.9993	.9992	.9991	.9989
20	.9999	.9999	.9998	.9998	.9998	.9998	.9997	.9997	.9996	.9996
21	1.0000	1.0000	.9999	.9999	.9999	.9999	.9999	.9999	.9998	.9998
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9999

x \ λ	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10.0
0	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0000
1	.0011	.0010	.0009	.0009	.0008	.0007	.0007	.0006	.0005	.0005
2	.0058	.0053	.0049	.0045	.0042	.0038	.0035	.0033	.0030	.0028
3	.0198	.0184	.0172	.0160	.0149	.0138	.0129	.0120	.0111	.0103
4	.0517	.0486	.0456	.0429	.0403	.0378	.0355	.0333	.0312	.0293
5	.1098	.1041	.0986	.0935	.0885	.0838	.0793	.0750	.0710	.0671
6	.1978	.1892	.1808	.1727	.1649	.1574	.1502	.1433	.1366	.1301
7	.3123	.3010	.2900	.2792	.2687	.2584	.2485	.2388	.2294	.2202
8	.4426	.4296	.4168	.4042	.3918	.3798	.3676	.3558	.3442	.3328
9	.5742	.5611	.5479	.5349	.5218	.5089	.4960	.4832	.4705	.4579
10	.6941	.6820	.6699	.6576	.6453	.6329	.6205	.6080	.5955	.5830
11	.7932	.8732	.7730	.7626	.7520	.7412	.7303	.7193	.7081	.6968
12	.8684	.8607	.8529	.8448	.8364	.8279	.8191	.8101	.8009	.7916
13	.9210	.9156	.9100	.9042	.8981	.8919	.8853	.8786	.8716	.8645
14	.9552	.9517	.9480	.9441	.9400	.9357	.9312	.9265	.9216	.9165
15	.9760	.9738	.9715	.9691	.9665	.9638	.9609	.9579	.9546	.9513
16	.9878	.9865	.9852	.9838	.9823	.9806	.9789	.9770	.9751	.9730
17	.9941	.9934	.9927	.9919	.9911	.9902	.9892	.9881	.9870	.9857
18	.9973	.9969	.9966	.9962	.9957	.9952	.9947	.9941	.9935	.9928
19	.9988	.9986	.9985	.9983	.9980	.9978	.9975	.9972	.9969	.9965
20	.9995	.9994	.9993	.9992	.9991	.9990	.9989	.9987	.9986	.9984
21	.9998	.9998	.9997	.9997	.9996	.9996	.9995	.9995	.9994	.9993

x \ λ	9.1	9.2	9/3	9.4	9.5	9·6	9.7	9.8	9.9	10.0
22	.9999	.9999	.9999	.9999	.9999	.9998	.9998	.9998	.9997	.9997
23	1.0000	1.0000	1.0000	1.0000	.9999	.9999	.9999	.9999	.9999	.9999
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x \ λ	11	12	13	14	15	16	17	18	19	20
0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2	.0012	.0005	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000
3	.0049	.0023	.0011	.0005	.0002	.0001	.0000	.0000	.0000	.0000
4	.0151	.0076	.0037	.0018	.0009	.0004	.0002	.0001	.0000	.0000
5	.0375	.0203	.0107	.0055	.0028	.0014	.0007	.0003	.0002	.0001
6	.0786	.0458	.0259	.0142	.0076	.0040	.0021	.0010	.0005	.0003
7	.1432	.0895	.0540	.0316	.0180	.0100	.0054	.0029	.0015	.0008
8	.2320	.1550	.0998	.0621	.0374	.0220	.0126	.0071	.0039	.0021
9	.3405	.2424	.1658	.1094	.0699	.0433	.0261	.0154	.0089	.0050
10	.4599	.3472	.2517	.1757	.1185	.0774	.0491	.0304	.0183	.0108
11	.5793	.4616	.3532	.2600	.1848	.1270	.0847	.0549	.0347	.0214
12	.6887	.5760	.4631	.3585	.2676	.1931	.1350	.0917	.0606	.0390
13	.7813	.6815	.5730	.4644	.3632	.2745	.2009	.1426	.0984	.0661
14	.8540	.7720	.6751	.5704	.4657	.3675	.2808	.2081	.1497	.1049
15	.9074	.8444	.7636	.6694	.5681	.4667	.3715	.2867	.2148	.1565
16	.9441	.8987	.8355	.7559	.6641	.5660	.4677	.3751	.2920	.2211
17	.9678	.9370	.8905	.8272	.7489	.6593	.5640	.4686	.3784	.2970
18	.9823	.9626	.9302	.8826	.8195	.7423	.6550	.5622	.4695	.3814
19	.9907	.9787	.9573	.9235	.8752	.8122	.7363	.6509	.5606	.4703
20	.9953	.9884	.9750	.9521	.9170	.8682	.8055	.7307	.6472	.5591

21	.9977	.9939	.9859	.9712	.9469	.9108	.8615	.7991	.7255	.6437
22	.9990	.9970	.9924	.9833	.9673	.9418	.9047	.8551	.7931	.7206
23	.9995	.9985	.9960	.9907	.9805	.9633	.9367	.8989	.8490	.7875
24	.9998	.9993	.9980	.9950	.9888	.9777	.9594	.9317	.8933	.8432
25	.9999	.9997	.9990	.9974	.9938	.9869	.9748	.9554	.9269	.8878
26	1.0000	.9999	.9995	.9987	.9967	.9925	.9848	.9718	.9514	.9221
27	1.0000	.9999	.9998	.9994	.9983	.9959	.9912	.9827	.9687	.9475
28	1.0000	1.0000	.9999	.9997	.9991	.9978	.9950	.9897	.9805	.9657
29	1.0000	1.0000	1.0000	.9999	.9996	.9989	.9973	.9941	.9881	.9782
30	1.0000	1.0000	1.0000	.9999	.9998	.9994	.9986	.9967	.9930	.9865
31	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9993	.9982	.9960	.9919
32	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9996	.9990	.9978	.9953
33	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9998	.9995	.9988	.9973
34	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9998	.9994	.9985
35	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9992
36	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9998	.9996
37	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9998
38	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999
39	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999

*The author would like to thank Mr. Morty Yalovsky for computing this table, and the McGill University, Department of Computer Science for computer facilities.

TABLE 4*

NORMAL PROBABILITY TABLE

Entry : $\int_0^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$

$x \backslash \lambda$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1627	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2882	.2910	.2939	.2967	.2996	.3023	.3051	.3079	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3290	.3315	.3340	.3365	.3389
1.0	.3414	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621

$x \backslash \lambda$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3906	.3925	.3943	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4146	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4278	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4453	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4648	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4762	.4767
2.0	.4773	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4865	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4914	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4933	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4986	.4987	.4987	.4988	.4988	.4988	.4989	.4989	.4990	.4990

*The author would like to thank Mr. Morty Yalovksy for computing this table, and the McGill University, Department of Computer Science for computer facilities.

TABLE 5*

STUDENT-t TABLE

Entry : $t_{a,v}$ where $\int_{t_{a,v}}^{\infty} f(t_v)dt_v = \alpha$ and $f(t_v)$

is the density function of a Student-t with y degrees of freedom

v	$a = .10$	$a = .05$	$a = .025$	$a = .01$	$a = .005$	v
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10

v	$a=.10$	$a=.05$	$a=.025$	$a=.01$	$a=.005$	v
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
inf.	1.282	1.645	1.960	2.326	2.576	inf.

*This table is abridged from Table IV of R. A. Fisher, *Statistical Methods for Research Workers*, published by Oliver and Boyd, Ltd., Edinburgh, by permission of the author and publishers.

TABLE 6*

THE χ^2 TABLE

Entry : $\chi^2_{a,v}$ where $\int_{\chi^2_{a,v}}^{\infty} f(\chi^2_v)d\chi^2_v = a$ and $f(\chi^2_v)$ is the density function of a chi-square variable with v degrees of freedom.

a	.995	.99	.975	.95	.10	.05	.025	.01	.005	.001
$v=1$	0.0^4393	0.0^3157	0.0^3982	0.00393	2.71	3.84	5.02	6.63	7.88	10.83
2	0.0100	0.0201	0.0506	0.103	4.61	5.99	7.38	9.21	10.60	13.81
3	0.0717	0.115	0.216	0.352	6.25	7.81	9.35	11.34	12.84	16.27
4	0.207	0.297	0.484	0.711	7.78	9.49	11.14	13.28	14.86	18.47
5	0.412	0.554	0.831	1.15	9.24	11.07	12.83	15.09	16.75	20.52
6	0.676	0.872	1.24	1.64	10.64	12.59	14.45	16.81	18.55	22.46
7	0.989	1.24	1.69	2.17	12.02	14.07	16.01	18.48	20.28	24.32
8	1.34	1.65	2.18	2.73	13.36	15.51	17.53	20.09	21.95	26.12
9	1.73	2.09	2.70	3.33	14.68	16.92	19.02	21.67	23.59	27.88
10	2.16	2.56	3.25	3.94	15.99	18.31	20.48	23.21	25.19	29.59
11	2.60	3.05	3.82	4.57	17.28	19.68	21.92	24.73	26.76	31.26
12	3.07	3.57	4.40	5.23	18.55	21.03	23.34	26.22	28.30	32.91
13	3.57	4.11	5.01	5.89	19.81	22.36	24.74	27.69	29.82	34.53
14	4.07	4.66	5.63	6.57	21.06	23.68	26.12	29.14	31.32	36.12

a	·995	.99	.975	·95	.10	.05	.025	.01	.005	.001
15	4.60	5.23	6.26	7.26	22.31	25.00	27.49	30.58	32.80	37.70
16	5.14	5.81	6.91	7.96	23.54	26.30	28.85	32.00	34.27	39.25
17	5.70	6.41	7.56	8.67	24.77	27.59	30.19	33.41	35.72	40.79
18	6.26	7.01	8.23	9.39	25.99	28.87	31.53	34.81	37.16	42.31
19	6.84	7.63	8.91	10.12	27.20	30.14	32.85	36.19	38.58	43.82
20	7.43	8.26	9.59	10.85	28.41	31.41	34.17	37.57	40.00	45.31
21	8.03	8.90	10.28	11.59	29.62	32.67	35.48	38.93	41.40	46.80
22	8.64	9.54	10.98	12.34	30.81	33.92	36.78	40.29	42.80	48.27
23	9.26	10.20	11.69	13.09	32.01	35.17	38.08	41.64	44.18	49.73
24	9.89	10.86	12.40	13.85	33.20	36.42	39.36	42.98	45.56	51.18
25	10.52	11.52	13.12	14.61	34.38	37.65	40.65	44.31	46.93	52.62
26	11.16	12.20	13.84	15.38	35.56	38.89	41.92	45.64	48.29	54.05
27	11.81	12.88	14.57	16.15	36.74	40.11	43.19	46.96	49.64	55.48
28	12.46	13.56	15.31	16.93	37.92	41.34	44.46	48.28	50.99	56.89
29	13.12	14.26	16.05	17.71	39.09	42.56	45.72	49.59	52.34	58.30
30	13.79	14.95	16.79	18.49	40.26	43.77	46.98	50.89	53.67	59.70
40	20.71	22.16	24.43	26.51	51.81	55.76	59.34	63.69	66.77	73.40
50	27.99	29.71	32.36	34.76	63.17	67.50	71.42	76.15	79.49	86.66
60	35.53	37.48	40.48	43.19	74.40	79.08	83.30	88.38	91.95	99.61
70	43.28	45.44	48.76	51.74	85.53	90.53	95.02	100.4	104.2	112.3
80	51.17	53.54	57.15	60.39	96.58	101.9	106.6	112.3	116.3	124.8
90	59.20	61.75	65.75	69.13	107.6	113.1	118.1	124.1	128.3	137.2
100	67.33	70.06	74.22	77.93	118.5	124.3	129.6	135.8	140.2	149.4

*This table is taken from Cambridge Elementary Statistical Tables, D. V. Lindley and J. C. P. Miller by permission of the authors and the publishers.

TABLE 7 (a)*

5% POINT OF F-DISTRIBUTION

$$\text{Entry : } F_{.05,\, v1,\, v2} \int_{F_{.05,\, v1,\, v4}}^{\infty} f(F_{v1,\, v2})\, dF_{v1,\, v2} = 0{\cdot}05 \text{ and } f(F_{v1,\, v2})$$

is the density function of a F-variable with $_{v1}$ and $_{v2}$ degrees of freedom.

$v_1=$	1	2	3	4	5	6	7	8	10	12	24	00
$v_2=1$	161.4	199.5	215.7	224.6	230.4	234.0	236.8	238.9	241.9	243.9	249.0	254.3
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.5	19.5
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.79	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.96	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.74	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.06	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.37	3.64	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.35	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.38	3.37	3.29	3.23	3.14	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.98	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.85	2.79	2.61	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.75	2.69	2.51	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.67	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.60	2.53	2.35	2.13

$v_1=$	1	2	3	4	5	6	7	8	10	12	24	∞*
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.54	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.49	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.45	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.41	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.38	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.35	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.32	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.30	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.27	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.25	2.18	1.98	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.24	2.16	1.96	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.22	2.15	1.95	1.69
27	4.21	3.25	2.96	2.73	2.57	2.46	2.37	2.31	2.20	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.19	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.18	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.16	2.09	1.89	1.62
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.14	2.07	1.86	1.59
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.12	2.05	1.84	1.57
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.11	2.03	1.82	1.55
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.09	2.02	1.81	1.53
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.08	2.00	1.79	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	1.99	1.92	1.70	1.39
100	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.91	1.83	1.61	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.83	1.75	1.52	1.00

*This table is taken from Cambridge Elementary Statistical Tables, D. V. Lindley and J.C. P. Miller by permission of the authors and the publishers.

TABLE 7 (b)*

THE F-DISTRIBUTION

Entry $F_{.01, v1, v2}$ where $\int_{F_{01, v1, v2}}^{\infty} f(F_{v1, v2})\, d\, Fv_1, v_2 = 0{\cdot}01$ and $f(F_{v1, v2})$

is the density function of an F with $_{v1}$ and $_{v2}$ degrees of freedoms.

$v =$	1	2	3	4	5	6	7	8	10	12	24	00
V =1	4052	50.0	5403	5625	5764	5859	5928	5981	6056	6106	6235	6366
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.2	27.1	26.6	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.5	14.4	13.9	13.5
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.05	9.89	9.47	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.87	7.72	7.31	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.62	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.81	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.26	5.11	4.73	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.85	4.71	4.33	3.91
11	0.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.54	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.30	4.16	3.78	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	.444	4.30	4.10	3.96	3.59	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	3.94	3.80	3.43	3.00

$v=$	1	2	3	4	5	6	7	8	10	12	24	∞
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.80	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.69	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.59	3.46	3.08	2.65
18	8.20	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.51	3.37	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.43	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.37	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.31	3.17	2.80	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.26	3.12	2.75	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.21	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.17	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.13	2.99	2.62	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.09	2.96	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.06	2.93	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.03	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.00	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	2.98	2.84	2.47	2.01
32	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	2.93	2.80	2.42	1.96
34	7.45	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.90	2.76	2.38	1.91
36	7.40	5.25	4.38	3.89	3.58	3.35	3.18	3.05	2.86	2.72	2.35	1.87
38	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.83	2.69	2.32	1.84
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.80	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.63	2.50	2.12	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.47	2.34	1.95	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.32	2.18	1.79	1.00

*This table is taken from Cambridge Elementary Statistical Tables, D.V. Lindley and J. C. P. Miller by permission of the authors and the publishers.

ANSWERS FOR SELECTED QUESTIONS

Chapter 1

1.3. (1, 1), (2, 1),......(6, 1)

(1, 2), (2, 2),......(6, 2)

....

(1, 6), (2, 6),......(6,6)

1.4. (*a*) $\{1, 2, \ldots\ldots, 52\}$ where the cards are numbered from 1 to 52.

(*b*) $\{(x, y \mid x, y \in \{1, 2, \ldots, 52\}\}$

(*c*) $\{(x, y) \mid x, y \in \{1, 2, \ldots, 52\}, x \neq y\}$

1.5. $\{1, 2, \ldots, 14\}$ (the balls are numbered from 2 to 14).

1.6. $\{(a, b) \mid a \in \{1, 2, \ldots, 52\}, b \in \{0, 1, -1\}, b=-1 \mid a \in \{27, 28, \ldots, 52\}, b \in \{0, 1\} \mid a \in \{1, 2, \ldots, 26\}\}$. (The cards are numbered from 1 to 52 of which 1 to 26 are red cards. The faces of the coin are assigned the numbers 0 and 1 if no coin is thrown the outcome is denoted by -1).

1.7. (*a*) 35 sets, (*b*) 840 vectors.

1.8. (*a*) $\{(0, 0, 1), (0, 1, 0), (1, 0, 0)\}$; (*b*) same as in (*a*) ;

(*c*) $\{(0, 1, 1), (1, 0, 1)\ (1, 1, 0), (1, 1, 1)\}$. Heads and tails are denoted by 1 and 0 respectively.

1.9. $\{2, 4, 6\}$

1.10. $V_1+V_2=(-4, 8, 5)$; $V_1V_2'=7$

1.11. $(6, -5, 0, 8)$ (not unique)

1.12. $V_1=(0, 1, 1)$; $V_2=(2, -1, 1)$ (not unique)

1.13. Not linearly independent.

1.16.
$$\begin{bmatrix} 1 & 1 & 3 & 4 \\ 2 & 1 & 0 & 1 \\ 0 & -3 & 5 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 5 \end{bmatrix}$$

1.17. (*a*) $\begin{bmatrix} 1/2 & 1/2 \\ 1 & 0 \end{bmatrix}$, $\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & 0 & 1/3 \\ 1/2 & 1/2 & 0 \end{bmatrix}$

(*b*) $\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$, $\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & 0 & 1/3 \\ 0 & 2/3 & 1/3 \end{bmatrix}$

1.19. $\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$

1.24. $B = \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix}$

1.27. 316.

Chapter 2

2.1. (*a*) 64 ; 15 excluding ϕ,
(*b*) 15 ; 7 excluding ϕ.

2.2. 18. **2.3.** 25.26^3.

2.4. (1) 24 ; (2) 6 ; (3) 360. **2.5.** 25200.

2.6. (*a*) 30 ; (*b*) 100. **2.7.** (*a*) $\binom{30}{5}$; (*b*) $\binom{29}{4}$.

2.9. 6^3. **2.10.** $\binom{13}{4}\binom{13}{5}\binom{13}{4}\binom{13}{1}$

2.11. (*a*) 120 ; (*b*) 80. **2.12.** $\binom{80}{16}\binom{20}{4}$.

2.14. (*d*) 3/8 ; $-45/2048$. **2.16.** 60.

2.17. not unique. $A=\{2, 5\}$, $B=\{2, 0, -1\}$; $A=\{2\}$, $B=\{2, 5, 0, -1\}$.

2.18. (*a*) Events of getting exactly 0 head, 1 head and 2 heads respectively.
(*b*) {(T, T, T), (H, H, H), (H, H, T), H, T, H), (T, H, H)}.

2.19. (1) {4, 5, 6}, (2) {0, 1, 2, 3, 4, 5}, (3) {3}, (4) {0, 1, 2}.

2.21. (*a*) yes, (*b*) no, (*c*) no, (*d*) no.

2.23. 0.7 ; 0.8 ; 0.7 ; 0.8 ; 0.3 ; 0.2 ; 0.3.

2.24. 2^8. **2.25.** 3/5.

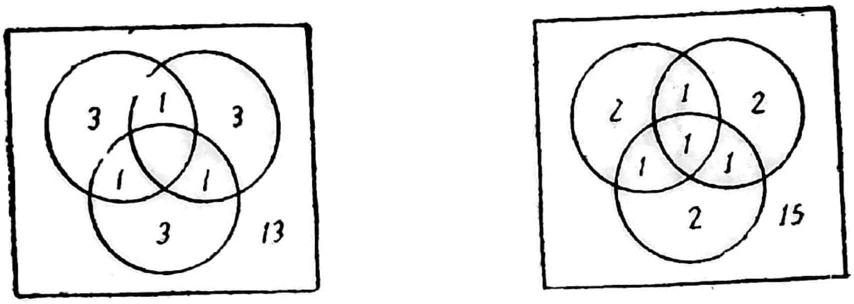
2.27. 15/36. **2.28.** $\binom{90}{15}\binom{10}{5}\Big/\binom{100}{20}$.

2.29. $\binom{200}{20}\binom{100}{10}\binom{50}{10}\binom{50}{0}\Big/\binom{400}{40}$.

2.30. yes. **2.31.** (*a*) 1/16, (*b*) 3/51.

2.32. Consider an experiment of throwing a coin twice. (*a*) Events of getting 0 head and 3 heads, (*b*) 0 head and one head, (*c*) {(T, T), (T, H)} and {(T, T), (H, T)}, (*d*) exactly one head and at least one head.

2.33.



2.34. 1/3. **2.35.** 2/5.

2.36. (*a*) 0.80, (*b*) 0, (*c*) 0 20. **2.37.** 3/8.

2.38. (*a*) 2/7, (*b*) 5/7. **2.39.** 2/3.

2.40. 2/3. **2.41.** 0·65.

2.42. 1/23. **2.43.** 0·5558.

2.44. 0 5558.

Chapter 3

3.1. X-the number of heads in the outcomes ; Y—2 times the number of tails in the outcomes ; (*b*) X-sum rolled, Y-difference rolled ; (*c*) X-number of boys in the outcomes, Y—2 times the number of girls —3, in the outcomes.

3.2. $f(x)=\binom{4}{x}\binom{48}{13-x}\Big/\binom{52}{13}$, $x=0, 1, 2, 3, 4.$

$$F(x')=\sum_{x=0}^{x'} f(x).$$

3.3. $f(x)=\binom{50}{x}(0.01)^x(0.99)^{50-x}$, $F(x')=\sum_{x=0}^{x'} f(x)$.

3.4. $f(x)=\binom{100}{x}(1/2)^x(1/2)^{100-x}$.

3.5. $f(x)=\binom{10}{x}\binom{20}{12-x}\Big/\binom{30}{12}$.

3.6. (*a*) no, (*b*) yes, $F(x)=\begin{cases}0, & x<-1\\ 1/3, & -1\leqslant x<0\\ 2/3, & 0\leqslant x<5\\ 1, & x\geqslant 5.\end{cases}$

(*c*) no, (*d*) yes, $F(x)=\begin{cases}0, & x<0\\ x^2/2, & 0\leqslant x<1\\ 2x-x^2/2-1, & 1\leqslant x<2\\ 1, & x\geqslant 2.\end{cases}$

3.8. (*a*) $k=1$, (*b*) 3/2, (*c*) 1/2.

3.9. (*a*) 1/8, (*b*) 7/16, (*c*) 5/8.

3.10. (*a*) 1/2, (*b*) 1/4, (*c*) 1/2, (*d*) 3/4.

3.11. $f(x)=\begin{cases}4x/5, & 0<x<1\\ 2(3-x)/5, & 1\leqslant x<2\\ 0 & \text{elsewhere.}\end{cases}$

3.12. (1) $F(x)=\begin{cases}0, & x<-2\\ 1/8, & -2\leqslant x<-1\\ 3/8, & -1\leqslant x<0\\ 6/8, & 0\leqslant x<2\\ 1, & x\geqslant 2.\end{cases}$; (2) 3/8 ; (3) 5/8

3.15. 50. **3.16.** \$ 1.

3.17. 2·02. **3.18.** 2·23.

3.20. (1) 0 ; $\sqrt{7/2}$, (2) $1/\theta$; $1/\theta$.

3.24. (*a*) Binomial, $n=10$, $p=1/2$

(*b*) Binomial, $n=4$, $p=2/3$

(*c*) Binomial, $n=3$, $p=3/5$.

3.25. $(e^{2t}-1)/t^2$.

3.27. (1) $\theta/2$; (2) $0{\cdot}2\theta$; (3) $\theta/2$; (4) θ ; (5) $\theta/2$; (6) $\theta/\sqrt{12}$.

3·29. (1) 2 ; (2) $\sqrt{4.8}$.

3·30. (1) 9/10 ; (2) 39/40 ; (3) 89/90.

3.31. (1) 3/8 ; (2) 1/4 ; (3) $\alpha/4$.

Chapter 4

4.1. (*a*) 0.1631 ; (*b*) 0.9692 ; (*c*) 0.3087.

4·2. (*a*) 0.2344 ; (*b*) 0·6563 , (*c*) 0.6562.

4.3. 0 0148. **4·5.** 0.0881.

4·6. $10/3^5=0.0412$. **4.8.** (*a*) 0.1513 ; (*b*) 0·5897.

4.9. (*a*) 0.0805 , (*b*) 0.8110.

4.10. 0.00513, 0.0071 ; 0.0285, 0.0164 ; 0.0775, 0.0409 ; 0.1380, 0.0683 ; 0.1800, 0.0854 ; 0.1843, 0.0854.

4.12. 0.00096, 0 00106.

4.13. (*a*) Binomial, N=5, $p=1/2$;
(*b*) Binomial, N=3, $p=1/3$ (*c*) Poisson, $\lambda=2$.

4.14. λ ; λ.

4.15. (*a*) 0.08854 ; (*b*) 0.2438.

4.16. (*a*) 0.0323 ; (*b*) 0.000.

4.17. (*a*) 0.0000 ; (*b*) 0.9706.

4.18. (*a*) $\sum_{i=1}^{n} e^{tx_i}/n$; (*b*) $pe^t(1-qe^t)^{-1}$, $q=1-p$.

(1) Geometric, $p=1/5$; (2) Geometric, $p=1/2$.

4.19. 0.0000.

4.20. (*a*) 0.0040 ; (*b*) 0.1005.

4.21. 0.00001.

4.22. $-(1-p)/p \log p$.

4.24. (*a*) $(e^{t\beta}-e^{t\alpha})/t(\beta-\alpha)$; (*b*) $(1-\theta t)^{-1}$.

4.25. (*a*) Gamma, $\alpha=2$, $\beta=3$; (*b*) Exponential, $\theta=1$.

4.31. (1) e^{-2} ; (2) $2e^{-1}$.

4.32. $e^{-3/2}$; $e^{-1/2}$.

4.33. e^{-125}.

4.34. $122e^{-10}$.

4.35. (*a*) $(1/2)\, e^{\beta^2/2}$; (*b*) $e^{\beta^2}\,(2\alpha\, e^{\beta^2}-1)/4$.

4.36. (*a*) $ap/(p-1)$; (*b*) $a^2p/(p-1)^2(p-2)$.

4.38. (*a*) 0.999 ; (*b*) 0.1062 ; (*c*) 0.9938 ;
(*d*) 0.0000 ; (*e*) 0.0000.

4.39. (*a*) 1.96 ; (*b*) 2.58.

4.40. (*a*) 1.96 ; (*b*) no.

4.41. 0 9772 ; 70.66.

4.42. (1.9804, 2.0196).

4.43. (1) 0.9772 ; (2) 0.3811.

4.44. 0.0456 ; <0.25.

4.48. $f(x)=\sqrt{2/\pi}\; e^{-x^2/2}$, $0<x<\infty$ and 0 elsewhere.

4.50.
$$f(y)=\begin{cases} 1/3, & 1<y<2 \\ (1/\sqrt{y-1}-1/3)/2, & 2\leqslant y<10 \\ 0 & \text{elsewhere.} \end{cases}$$

4.51. (*a*) $N(\mu=0,\ \sigma=1)$; (*b*) Gamma, $\alpha=2$, $\beta=3/2$;
(*c*) Exponential, $\theta=5$; (*d*) Poisson, $\lambda=2$.

4.52. (*a*) Binomial, $p=2/3$, N=5 ; (*b*) Geometric ; (*c*) Rectangular;
(*d*) $N(\mu=2,\ \sigma=\sqrt{2})$.

Chapter 5

5.1. (1) X —number of red marbles in the outcomes ; Y—number of white marbles ; $f(0,\ 3)=3^3/8^3$; $f(1,\ 2)=3^2.5^2/8^3$; $f(2\ ;\ 1)=3^2.5/8^3$; $f(3,0)=5^3/8^3$ and 0 elsewhere.

(2) X —number of red marbles, Y —number (red-white) ; $f(0, -3)=3^3/8^3$, $f(1, -1)=3^2.5^2/8^3$, $f(2, 1)=3^3.5/8^3$, $f(2, 3)=5^3/8^3$.

5.2. (1) $f(2, 0)=1/36$, $f(3, -1)=1/36$...., $f(12, 0)=1/36$.

(2) $F(2, 0)=1/36$, $F(3, -1)=2/36$,..., $F(x, y)=1$ for $x \geqslant 12$, $y \geqslant 0$.

(3) $f(1)=1/6$. $f(2)=1/6$,..., $f(6)=1/6$ and $f(x)=0$ elsewhere $g(1)=1/6$, $g(2)=1/6$,..., $g(6)=1/6$ and $g(x)=0$ elsewhere.

(4) $f(1 \mid y=4)=1/6$,..., $f(6 \mid y=4)=1/6$ and $f(x|y)=0$ elsewhere.

5.3. (1) $F(-1, 0)=1/15$, $F(-1, 1)=4/15$, $F(-1, 2)=6/15$

$F(0, 0)=3/15$, $F(0, 1)=8/15$, $F(0, 2)=11/15$

$F(1, 0)=4/15$, $F(1, 1)=10/15$, $F(1, 2)=15/15$

and $F(x, y)=1$ for $x \geqslant 1$. $y \geqslant 2$.

(2)
$$f(x)=\begin{cases} 6/15, & x=-1 \\ 5/15, & x=0 \\ 4/15, & x=1 \\ 0 & \text{elsewhere.} \end{cases}$$

(3)
$$f(x \mid y=2)=\begin{cases} 2/5, & x=-1 \\ 1/5, & x=0 \\ 2/5, & x=2 \text{ and } 0 \text{ elsewhere.} \end{cases}$$

5.4.
$$f(x)=\begin{cases} 3/12, & x=0 \\ 9/12, & x=1 \\ 0 & \text{elsewhere} \end{cases} \qquad g(y)=\begin{cases} 3/12, & y=0 \\ 3/12, & y=1 \\ 6/12, & y=2 \\ 0 & \text{elsewhere.} \end{cases}$$

$$f(y \mid x=0)=\begin{cases} 1/3, & y=0 \\ 1/3, & y=1 \\ 1/3, & y=2 \text{ and } 0 \text{ elsewhere.} \end{cases}$$

5.5. (1) $8/105$; (2) $2(x+1)/3$, $0<x<1$; (3) $2(z+3)/7$, $0<z<1$.

5.7. (1) $6/15$; (b) $e^{-1/2}$. **5.8.** $15/256$.

5.9. $1/2-5e^{-2}/6$. **5.10.** $9/10-e^{-1}+e^{-10}/10$.

5 11. (1)1 ; (2) 1/2 ; (3) 0 ; (4) 0.

5.12. (1) 2/9 ; (2) 1/9 ; (3) 0 ; (4) 0.

5.13. 0.

5.19. 3/7.

5.21. $(p_1e^{t_1}+\ldots+p_ke^{t_k})^N$; Np_i ; $N(N-1)p_i^2$; $N(N-1)p_j^2$; $N(N-1)p_ip_{ti}$.

5.22. 0·8994.

5.24. $\mu_1=10$; $\mu_2=15$; $\sigma_1=2$; $\sigma_2=3$; $\rho=1/2$.

5.25. (1) 0·9109 ; (2) $1-e^{-3/2}$.

5.27. (*a*) yes ; (*b*) yes.

5.28. (1) $\mu_1+2\mu_2$, $\sigma_1^2+4\sigma_2^2$, $\sigma_1^2+4\sigma_2^3+4 \text{ cov } (x, y)$.

(2) $\mu_1-\mu_2$, $\sigma_1^2+\sigma_2^2$, $\sigma_1^2+\sigma_2^2-2 \text{ cov } (x, y)$.

5.29. Var(X)—Var(Y).

5.30. Bivariate, $\mu_1=0=\mu_2$, $\sigma_1=\sigma_2=1$, $\rho=0$.

5.33. $N(\mu, \sigma/\sqrt{n})$.

Chapter 6

6.3. (1) Binomial, $N=55$, $p=1/2$; (2) Poisson, $\lambda=11$.

6.4. $f(y)=\begin{cases} 4y/\theta^2, & 0<y<\theta/2 \\ 4(\theta-y)/\theta^2, & \theta/2<y<\theta \end{cases}$ and 0 elsewhere.

6.5. Gamma, $\beta=3$, $\alpha=n(n+1)a/2$.

6.6. (*a*) $N(\mu_1-2\mu_2+\mu_3, \sqrt{\sigma_1^2+4\sigma_2^2+\sigma_3^2})$;

(*b*) $N(\mu_1+\mu_2-5\mu_3, \sqrt{\sigma_1^2+\sigma_2^2+25\sigma_3^2})$

6.7. $\sum_{i=1}^{N} p_i$ $\quad \sum_{i=1}^{N} p_i(1-p_i)$

6.8. $\sum_{n=0}^{N} \binom{N}{n} p^n(1-p)^{N-n} . (n/2\pi)^{1/2} e^{-n(x-\mu)^2/2\sigma^2}$.

6.9. (1) μNp ; (2) $Np\sigma^2+\mu^2Np(1-p)$.

6.10. $N(\mu, \sigma)$.

6.11. $\bar{x}=2.4$, $\bar{y}=2{\cdot}4$, $s_1^2=3.44$, $s_2^2=23.84$, $r=0.9767$.

6.12. (1) $X=(X_1+X_2)/2$, $Y=(Y_1+Y_2)/2$;

(2) $\sum_{i=1}^{2} (X_i-\bar{X})^2/2 \quad \sum_{i=1}^{2} (Y_i-\bar{Y})^2/2$; (3) $\sum_{i=1}^{2} (X_i-\bar{X})(Y_i-\bar{Y})/2$.

6.13. $(\sigma_1^2/n_1+\sigma_2^2/n_2+4\sigma_3^2/n_3)^{1/2}$.

6.14. $\sigma(1/n_1+1/n_2)^{1/2}$.

6.15. $\geqslant 0.7333$.

6.16. $\leqslant (16.6/15)^{1/2}$.

6.17. 1.

6.18. 0.0000.

6.19. $\geqslant 0.9756$.

6.20. $f(u_1)=n(\beta-u_1)^{n-1}/(\beta-\alpha)^n$, $\alpha<u_1<\beta$

$g(u_2)=n(u_n-\alpha)^{n-1}/(\beta-\alpha)^n$, $\alpha<u_n<\beta$.

6.21. $f(R)=n(n-1)R^{n-2}[(\beta-\alpha)-R]/(\beta-\alpha)^n$, $0<R<\beta-\alpha$.

6.22. $\dfrac{(2n+1)\,!}{(n\,!)^2}\left(\int_{-\infty}^{m} f(x)dx\right)^n\left(\int_{m}^{\infty} f(x)\, dx\right)^n f(m)$,

where $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty<m<\infty.$

6.23. $(2\pi\sigma^2)/8n.$

6.24. $(1/\theta)e^{-R/\theta}$, $0<R<\infty$ $(n=2)$.

6.28. (1) μ ; (2) σ^2/n.

6.34. (a) 0.025 ; (b) 0.115.

6.35. (1) 6 ; (2) 0.9664.

Chapter 7

7.1. 1 approx :

7.2. 0.0047.

7.3. 0.9266.

7.4. (33.37, 36.63) ; no.

7.5. (33.56, 51.44), $(-\infty, 50)$.

7.6. 0.9936.

7.7. $(-3.658, -0.342)$; no.

7.8. accept.

7.9. accept.

7.12. 0.99.

7.13. 0.5 approx.

7.14. (9.11, 42.6).

7.15. 1 approx.

7.16. 0.95 approx.

7.17. (9849.2, 10150.8).

7.18. yes at 98%.

7.19. $0 \; ; \; v^2 \frac{\Gamma\left(\frac{3}{2}\right).\Gamma\left(\frac{v}{2}-1\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{v}{2}\right)}$ for $v>2$.

7.20. yes at 95%.

7.21. 0.0001 approx.

7.22. 0.0287 approx.

7.23. 31.4.

7.24. 0.4 approx.

7.25. $\mu_r' = \left(\frac{m}{n}\right)^r \frac{\Gamma\left(\frac{m}{2}+r\right)\Gamma\left(\frac{n}{2}-r\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}$

if $2r<n$ and do not exist if $2r \geqslant n$.

7.27. 0.9 approx., $P(F_{14, 19} \leqslant 2.04)$.

Chapter 8

8.1. (80.2282, 85.7718).

8.2. (64.2621, 65.7379).

8.3. (9930.2619, 10069.7381).

8.4. (4.9883, 5.0317) ; (1) no ; (2) yes.

8.5. ($ 0.95, 1.05).

8.6. (26.8093, 73.1907).

8.7. (0.526, 0.574).

8.8. (6.5888, 13.4112).

8.9. $(-2.4231, 4.4231)$.

8.10. (18.5359, 21.4641).

8.11. $(-8.4552, -1.5048)$.

8.12. (0, 2599).

8.13. (412, 1588)

8.14. $(-4.62\%, 2.62\%)$.

8.15. $(-14.51\%, 4.51\%)$.

8.16. (0.155, 0.299).

8.17. (14.7, 75.2).

8.18. (2.38, 8.58).

8.19. (184.5, 673.5).

8.20. $(s'/[1 \pm z_{\alpha/2}/\sqrt{2n}])$.

Chapter 9

9.1. 0.1 ; 0.1. **9.2.** 5.

9.3. $\hat{\beta}=\bar{x}/\hat{\alpha}$ and $\hat{\alpha}$ is obtained numerically from

$$(\Sigma \log x_i - n \log \bar{x}) + n \log \alpha - n \frac{\partial}{\partial \alpha} \log \Gamma(\alpha) = 0.$$

9.4. \$ 7000. **9.5.** \$17.5.

9.6. $\hat{\alpha}=\min(x_1, \ldots, x_n)$ and $\hat{\beta}=\max(x_1, \ldots, x_n)$.

9.7 $\hat{\theta}=(2\bar{x}-1)/(1-\bar{x})$; $(-1+\sqrt{1-4n\,\Sigma \log x_i})/2$.

9.8. $2+X(X-1)/N(N-1)$. **9.14.** x/N.

9.15. θ_1 ; not both.

9.23. (1) $2\bar{X}$; (2) ΣX_i^2 ; (4) ΣX_i^2.

Chapter 10

10.1. $S=\{(G, G), (G, B), (B, G), (B, B)\}$; $C=\{(G, G)\}$.

10.2 $S=\{(x, y) \mid -\infty<x<\infty, -\infty<y<\infty\}$; $C=\{(x, y) \mid (x+y)/2>5\}$.

10.3. $\alpha=1/8$; $\beta=19/27$.

10.4. $C=\{x \mid x \geqslant 15\}$; $\alpha=e^{-3}$; $\beta=1-e^{-3/2}$.

10.5. 5.

10 6. reject H_0 if (*a*) $ns^2/\sigma_0^2 \geqslant \chi^2_{\alpha, n-1}$;

(*b*) $\bar{x} \geqslant G_\alpha$ where $P\{\text{gamma with } \alpha=n, \beta=\theta/n \text{ is } \geqslant G_\alpha\}=\alpha$;

(*c*) $x \geqslant k_\alpha$ where $\Sigma^n_{k_\alpha} \binom{n}{p_0} p_0^x (1-p_0)^{n-x} \geqslant \alpha$.

10.8. $\alpha=0.14$, $\beta=0.717$.

10.10. 0 ; 1/5 ; 16/45 ; 21/45 ; 24/45 ; 24/45 ; 21/45 ; 16/45 ; 1/5 ; 0.

10.18. (1) reject ; (2) reject. **10.20.** accept.

10.22. reject. **10.24.** reject.

10.26. accept the claim. **10.28.** 18.

10.30. reject. **10.32.** 0.7974 ; 0.4119 ; 0.0171.

10.34. yes. **10.36.** reject.

Chapter 11

11.2. no. **11.4.** not a good fit.

11.6. no.

11.8. There is evidence of independence ; $p=0.28$.

Chapter 12

12.2. (1) 1; (2) 2; (3) ∞.

12.4. (1) $a=-0.007$, $b=-0.475$, $c=-0.05$;
(2) $a=1.10$, $b=21.18$; (3) $a=4.86$, $b=0.106$.

12.6. (1) 8/15.

12.8. $a=\mu_1-b\mu_2-c\mu_3$.

$$b=\left(\sigma_{12}\ \sigma_3^2-\sigma_{23}\ \sigma_{13}\right)\Big/\left(\sigma_2^2\ \sigma_3^2-\sigma_{23}^2\right)$$

$$=\frac{\sigma_1}{\sigma_2}(\rho_{12}-\rho_{13}\ \rho_{22})\Big/\left(1-\rho_{23}^2\right).$$

$$c=\left(\sigma_{13}\ \sigma_2^2-\sigma_{23}\ \sigma_{12}\right)\Big/\left(\sigma_2^2\ \sigma_3^2-\sigma_{23}^2\right)$$

$$=\frac{\sigma_1}{\sigma_3}(\rho_{13}-\rho_{23}\ \rho_{12})\Big/\left(1-\rho_{23}^2\right).$$

12.10. $a=1430.0$, $b=-61.6$, $c=-11.4$.

12.12. $\Sigma(x_i-\mu)^2/n$.

12.14. 7.5.

12.20. significant

12.26. not significant (methods).

12.28. not significant (methods).

12.30. no. without more assumptions.

12.34. not good.

12.36. accept.

12.40. accept.

INDEX

A

B

C

D

RODUCTION TO STATISTICAL MATHEMATICS